青少年百科

QINGSHAUNIAN BAIKE

多媒体世界



国家新课程教学策略研究组 编写

一个虚拟的世界，一个声光编织的精彩世界，
一个令你为之疯狂陶醉的世界……

新疆青少年出版社
喀什维吾尔文出版社

青少年百科

qing shao nian bai ke

多媒体世界

国家新课程教学策略研究组/编写

新疆青少年出版社
喀什维吾尔文出版社

图书在版编目(CIP)数据

多媒体世界/国家新课程教学策略研究组编写. 修订本. 喀什:喀什维吾尔文出版社;乌鲁木齐:新疆青少年出版社,2007.10

(青少年百科)

ISBN 978－7－5373－1083－3

Ⅰ.多… Ⅱ.国… Ⅲ.多媒体技术－青少年读物 Ⅳ.TP37－49

中国版本图书馆CIP数据核字(2007)第170374号

青少年百科

多媒体世界

国家新课程教学策略研究组/编写

新疆青少年出版社
喀什维吾尔文出版社 出版

北京市朝教印刷厂印刷

787毫米×1092毫米 32开 480印张 14000千字

2007年11月修订版 2007年11月第1次印刷

印数:1－5000册

ISBN 978－7－5373－1083－3

前言

随着新课程改革浪潮的一步步推进，我国基础教育课程改革取得了令人欣喜的成就，基础教育课程改革，也是关系全社会的一件大事。

《基础教育课程改革纲要（实行）》中强调，为提高民族素质，增强综合国力，必须全面推进基础教育课程的改革。对中小学生进行素质教育的热潮正如火如荼的进行，并日益加快步伐。根据课程的资源和学生的需求，为了增加中小学生的课外阅读面，提高学生的阅读能力和全面素质的发展，我们组织了相关专家，编写了此套丛书。

丛书内容丰富、实用，深入浅出。选材时基本上是以知识性为标准的，但也兼顾到了可读性，可以说是知识性、可读性都很强的青少年读物。

由于编著水平有限和选择资料的工作量太大以及时间的关系，书中难免会出现一些疏漏、不当的地方，希望广大读者朋友能够理解，也欢迎给予批评指正。

编　者

目录

第一章　绪论 ………………………………………………… (1)
第一节　进入多媒体时代 ……………………………………… (1)
第二节　多媒体是指什么 ……………………………………… (5)
第三节　多媒体的各种媒体技术 ……………………………… (8)
第四节　多媒体技术的过去、现在和将来 …………………… (12)
第二章　历史悠久的计算机文本媒体 ……………………… (15)
第一节　字符编码 …………………………………………… (17)
第二节　文本输入 …………………………………………… (20)
第三节　文本输出 …………………………………………… (24)
第三章　形形色色、亦真亦幻的计算机图像媒体 ………………………………………………………… (28)
第一节　乐于助人的计算机图形 …………………………… (29)
第二节　神通广大的计算机图像 …………………………… (37)

第三节　寓教于乐的计算机动画 …………………… (46)
第四节　多彩多姿的计算机视频 …………………… (56)
第四章　能说会唱、善解人意的计算机声音媒体 ……
…………………………………………………… (64)
第一节　数字音频的一般处理 …………………… (66)
第二节　MIDI 合成音乐 …………………… (72)
第三节　语音的压缩编码 …………………… (74)
第四节　各种各样的语音处理技术 …………………… (81)
第五节　计算机声音媒体的广泛应用 …………………… (94)
第五章　五花八门的多媒体应用系统 …………………… (98)
第一节　多媒体辅助教学 …………………… (100)
第二节　多媒体电子出版业 …………………… (102)
第三节　多媒体通信系统 …………………… (112)
第四节　多媒体地理信息系统 …………………… (118)
第五节　多媒体监控系统 …………………… (121)

第一章　绪论

第一节　进入多媒体时代

进入 20 世纪 90 年代以来，不论是计算机界，还是一般计算机用户，谈论最多的莫过于多媒体和网络通信了，那么什么是多媒体呢？

对于现在的青少年朋友来说，电视已不是什么新鲜的事了，但是如果把日历往回翻上五六十年的话，恐怕就不会没人不对此感到好奇了。确实，电视的出现对于我们的工作和生活无疑是具有很大影响的，而就其对信息传播方式的影响来说则是“革命性”的，因为电视一改往日报纸、书籍、图片、广播等单纯以一种形式传播信息的手段，而是采用图、文、声多位一体的方式来

传播，这种想法本身就是一种创新和革命。事实上电视的出现不仅丰富了人们的日常生活，为人们提供了多方位的信息服务和视听享受，同时也带动了一大片工业的发展，使人类社会向前发展了一大步。但是随着社会的进步和发展，人们迎来了“信息爆炸”的时代，汹涌澎湃的信息浪潮铺天盖地地向我们涌来，时时冲击着我们传统的工作和生活方式。于是如何有效地处理和利用这些五花八门、良莠不齐的信息又成了时代摆在我们面前一道难题。显然面对如此浩瀚的信息海洋，一头扎进去是非“淹死”不可的，我们唯一的选择是选择一条好的“轮船”，即借助于信息处理工具，只有借助于它我们才能自由地畅游这烟波浩淼而又奇妙无穷的“信息海洋”。

那么这种信息处理工具应是怎么样的呢？电视因为具有综合传输图、文、声等形式信息的能力，无疑是一个比较好的“候选人”。但要命的是，电视虽具有传播信息的能力，但这种传播仅是单向的，用户只能被动地接收，却无法按自己的意愿主动地与之交互，另外电视对于复杂的信息加工竟也是一窍不通的。那么，我们只好另寻出路了。噢，对了，计算机不是号称“电脑”吗，应该具有这种信息处理能力吧？是的，计算机因为具有强大的运算能力，所以对于复杂的信息处理，它无疑是个合适的人选。但可惜的是，普通计算机仅能处理数字数据、文本数据和简单的图形数据，而对于复杂的图形、声音、图像、视频、动画等的处理竟表现出孩童般的无知。那么出路在哪里呢？

从上面的分析我们可以看出，电视和计算机对于图、文、声等信息的处理在功能上具有一定的互补性。电视具有丰富多

采、声像并茂的视频处理能力，但缺乏对此的灵活处理和复杂加工能力，并且也不具有交互能力；而计算机正好相反，不仅有灵活处理复杂信息的能力，交互性也好，且表现出一定的“智能”，只是目前在处理图、声方面有所欠缺。于是很自然的想法是把计算机和电视结合起来，取长补短充分发挥各自的优点。事实上，正是这种想法的结果直接导致了多媒体的出现。

当然一种理论和技术的出现并非是一蹴而就的，多媒体的概念也并非是一开始就马上提出来的，事实上多媒体技术最初是由计算机人机交互技术发展而来的。

大家知道计算机的出现延伸了人脑的一部分功能，成了人们现在最常使用的信息处理工具。不过计算机在社会中扮演的角色在很大程度上取决于它能处理的对象的性质。因为大家应该认识到，计算机并非是万能的、什么都能干的，其实计算机所做的工作一般是由人事先编好的，所以它对于重复性的劳动比较胜任，但对于诸如处理知识、计划、决策等大量颇具创造性的工作，却表现出十分的无能，尚需在人的参与下才能完成。于是计算机能否广泛应用的一个关键就在于人——计算机交互技术（简称人机交互或人机对话技术，这种技术，简单地说来就是研究计算机使用起来方不方便，符不符合人的习惯的问题）。大家知道人们交换信息时最习惯使用的方法莫过于眼看、耳听、手写，这就要求计算机在与人交互时，也能适应这几种信息交流方式，于是自然要求计算机必须具有处理图、文、声等多种信息形式的能力，这就促进了计算机多媒体技术的迅速发展。能综合处理图、文、声等多种形式信息既是多媒体技术的核心所在，也

是多媒体的魅力所在。

进入20世纪90年代以来,多媒体技术以前所未有的速度获得迅猛发展,它是目前世界上发展最快的产业之一,年平均增长率达到40%~50%,市场规模也由1992年的50亿美元发展到1996年的250亿美元。与此同时,与多媒体相关的可视化技术、虚拟现实技术也受到高度重视和发展。虚拟现实技术是一种模拟现实环境的技术,当人们与这种虚拟环境交互时,感觉犹如身临其境一般,可广泛应用于数字、娱乐、飞行器驾驶培训、军事应用等领域。

目前以MPEG为运动图像压缩技术标准的数字视频产品,如V——CD(Video——CD,视频光盘)等和HDTV(High——Definition Television,高清晰度电视)产品的发展最为引人注目。另外,以多媒体通信技术为基础的视频会议系统、个人可视通信系统也因大量高速通信网的建成和视频压缩处理技术的进步而变得越来越实用,而且通过全球卫星通信网络向全世界播送数字电视也已指日可待。而以CD——ROM光盘为代表的多种电子出版物、如电子游戏、光盘图书、光盘数据库、光盘杂志、电子报纸等更是早已深入了我们生活的每一方面,为我们的生活增色不少。

事实上,多媒体技术的发展,不仅带动了各种媒体处理技术的进步,也带动了一大片工业的发展,如光盘工业、电子出版业、游戏娱乐业、多媒体通信事业等,甚至连传统的电影电视业也从中得益匪浅,大家知道多媒体在影视业中的应用已不是什么新闻了,著名的影片如《终结者Ⅱ》、《侏罗纪公园》等无不用到了当

时最新的图形、图像和动画技术。多媒体的应用也从传统的教育、娱乐、信息服务等传统领域，扩展到工业、金融、国防、通信等更广泛的领域，甚至已深入到了生活的方方面面，如上街可看到各种各样的多媒体电子广告牌，购物有多媒体导购系统，宾馆有详细的宾馆信息咨询系统，银行也有多媒体银行远程服务系统可以免除我们的奔波之苦，甚至在街旁还能从多媒体信息咨询系统了解到最新的天气变化情况和股市行情等。怪不得有人说，20 世纪 90 年代以来，我们已经开始跨入多媒体信息时代了！

第二节　多媒体是指什么

处在电脑信息时代的我们，也许都听到过许多计算机方面的术语，相信“多媒体”就是其中之一。不过要想用一两句话解释清楚“多媒体”是什么，并不太容易，我们首先得知道计算机是如何与外界交换信息的。在此之前我们先谈谈人是如何与自然、与他人交换信息的。

信息对我们来说并不陌生，因为我们每天都在与它打交道，只是我们没有意识到而已。比如，我们早上听到公鸡打鸣，于是你就获得了这样一种信息——天快亮了；天亮起床后你看到许多人在插柳树，于是你又想到了：啊，春天来了，该植树了；但一

阵风吹来，你又不由得拉了拉衣服，哦，冬天的阴影还没完全消失嘛，天气还挺冷的。俗话说“落叶知秋”，就深刻地反映了我们从自然界获取信息的过程。

不仅如此，我们生活在社会中，还每时每刻地在与他人交换着信息。我们上学时都要与老师、同学交流，上班时要与领导、同事交往，上公园要与管理员、游人交往……另外除了直接与他人交往外，我们每天还要阅读大量的书报，收看很多的广播电视节目，处理数不清的账单发票……从而间接地与他人交流着信息。

所以信息与我们的生活息息相关，是不可分割的部分，以至于有人甚至认为信息同物质、能量一起构成了我们这个世界的三要素，其重要性可见一斑。但是信息又不同于物质、能量，因为它是无形的、抽象的东西，事实上，我们也只有通过物质才能获得信息，即信息的存在是需要用物质来体现的，或者说物质是信息的依托物或载体。对于间接的信息交流，这一点就显而易见了，因为我直接接触的本身就是书报、广播电视节目、账单发票等物质，而非信息本身。即使是人与人直接交流，我们采取的办法也无非是用语言口头表达或用文字、图等笔头写，这语声，这文字，这图不就是物质？当然，信息的传播也需要能量作为动力。

所以说，人们交流信息必须依赖于物质作为形式。事实上，信息也正是以物质作为载体的。信息的物质载体，在计算机科学中称为信息媒体。媒体即媒介物的意思，信息媒体的意思就是说信息需以物质为媒介物。人类能感知的信息媒体一般有两

种，这就是视觉系统能感知的图文，声觉系统感受到的声、言。

大家都知道计算机具有很强的运算能力。它的出现被认为是延伸了人脑的功能，这正如机械延伸了人手足的功能一样，有时计算机还被形象地称为电脑。确实，现代计算机在外部功能上与人脑有着相似之处，在应用系统中，它就常常作为整个系统的控制中心，这正如人脑是人体的神经中枢一样。既然计算机是控制中心，那就免不了与其他部分与人交换信息，这样就提出了计算机如何以某种方式与外界交换信息的问题。

计算机不管功能如何强(往往是在某些方面)，总免不了人的参与，而且人们发明计算机的目的就是为人服务嘛，所以就有一个计算机如何与人交流信息的问题，或称人机交互。而人输入信息最方便的手段是视听，输出信息最方便的就是说，稍慢一点的就是写。为了使人能方便地与计算机交互，就要求计算机也应能看能听能说甚至能唱，而且水平应更高，只有这样我们人用起来才方便。对于人，看、说、听、写的对象就是图、声、文等信息媒体，所以要求计算机至少也能处理这几种信息媒体，我们把这种能处理多种信息媒体的计算机就称为多媒体计算机(比如MPC，即多媒体个人计算机就是最常见的一种)。

多媒体计算机与传统计算机的区别就在于它不仅能处理一般抽象的数字符号，而且还能处理人们熟悉的图、文、声等多种信息媒体，所以确切地说多媒体计算机应称为多媒体信息处理机才比较合适。多媒体计算机的优点在于它不仅具有强大的信息处理能力，而且容易使用，使用起来符合人的习惯；不仅能说会唱，图文并茂，而且样样都是专业水平，令一般人望尘莫及。

当然这样的多媒体计算机还不能算是机器人，因为一般说来它尚不具备诸如学习、推理等智能，而且它也不具备“手足”，不能活动，当然在有些方面功能上与机器人也是具有一定的相似性的。

第三节　多媒体的各种媒体技术

前面一节已经提到了，多媒体计算机一般能处理图、文、声等信息媒体，一般分别被称为图像媒体、文本媒体和声音媒体。当然计算机能处理的信息媒体除此之外还有很多，比如超声波信号、地震信号、医用图像、遥感数据、红外线热像等，这些都应用在专门领域，涉及较多的专门知识，这里就不一一介绍了。而且事实上目前用的多媒体一般就指这上面三种，下面就对此作一简要介绍。

文本媒体是指由文字符号构成的信息媒体，这是在计算机上最早应用的信息媒体，它可用于记录源程序、文章、提示信息、构造数据库、文本表格等，在交互方式下还可以作为命令输入给计算机。文本最基本的是文字，如 26 个英文字母，中文的汉字，其他还有标点符号、数字数学符号等。对它的处理一般涉及输入、编辑（如重新编排和修理等）、输出，现在还有自动识别、语言理解等。

用文本媒体来储存一本小说也许是合适的方式，但对于一本画册，用文本来描述显然是力不从心了，这不仅表现在文字不直观，不生动，而且还表达不全，但如果用图像媒体则可以轻松地解决这个问题了。图像媒体简单说就是以图的方式（静止的或活动的）来储存信息，传递信息。图像媒体具体地说又包括下面四种：图形、静止图像、动画和视频，当然这四种并非完全不相容的，比如有的书就把一、三两个归为一类称为视频绘图，二、四归为一类称为视频。不过分类并不是最重要的，目的只是想说明它们之间的区别而已。

图形一般是指由计算机产生的静止图像，而动画是指由计算机产生的活动图像，但有的动画原型并不一定是计算机产生的，可由外面的实物通过摄影制成图像输入。静止图像和视频一般是由外界输入的，前者可由扫描仪摄像机等输入，后者则来源众多，电视机、摄像机、磁带、光盘等，两者区别就在于一个是静止画面，一个是活动画面。当然，前面的区别是相对的，因为它们往往是相互利用的，例如，静止图像就可用截取活动视频的方法获得。

当然图像媒体的这四个部分的区别还不止上面提到的这些，在处理方法、应用场合等方面也各有千秋。比如在计算机辅助设计中，常要求用户可以自由操作图中的每一部分，如挪动一个元器件的位置，或旋转 90°以利布线，或者缩小整个画面以便观察全局，或者放大某一部分以便修改等等，这就必须用到计算机图形，因为在这儿图中的每一部分都是计算机根据用户的要求生成的，当然可以自由改动。计算机动画也具有相应的灵活

性，但生成的是活动图像，因而更具生动性，常用于计算机辅助教学（如演示动态的物理过程，化学变化等），广告设计，电子游戏，以及影视片制作等。对静止图像的处理一般有压缩编码、恢复、增强、识别等，另外在屏幕上显示时，还可以有放大、缩小、旋转、移动、淡入淡出等各种视频特技（这在电视上是经常可以看到的）。电子相册，文档中的照片管理，图像识别（文字识别也是它的一种）等都是静止图像的用武之地，另外它还更多地应用于专门领域，如地震、勘探、医学等。视频也许是其中我们最熟悉的了，因为我们几乎每天都要观看电视节目，了解新闻、观看体育比赛、欣赏电视剧等。但计算机视频又与它有所不同，首先电视是模拟信号（即时间上和幅度上都连续的信号），而计算机视频则是数字信号（时间和幅度都不连续，即离散的），所以有时又称数字视频；其次由于数字信号有灵活性、便于处理的优点，这使我们的计算机视频往往是交互式的，双向的，而不像一般电视那样我们是被动接受，即电视台放什么，我们就只能看什么。因此我们在欣赏数字视频时，不仅可以选择节目，甚至还可以自由编排内容，安排情节进展，以至于扮演其中的一个角色。计算机视频的主要处理手法有施加各种特技、输入文字、合成各种音响效果和音乐等。

说到这里，我们已经提到音响和音乐等声音信号了，那么计算机是如何处理声音媒体的呢？

声音是由空气振动引起的，一般是随时间变化的，所以又称音频。但这儿我们又把它们粗分为三种：由人说话产生的语音，演奏乐器产生的音乐以及既有人声（如唱歌，和声等）又有音乐、

音响效果的音频。由于三种声音媒体各有不同之处，下面先作一简单介绍。

语音，即话音，是由人说话产生的，一般说来带宽较窄，而且常与语言融合在一起，通常含有特殊含义。对它的处理一般也涉及语义这一点，如语音编码和合成，语音识别，语音增强等，都围绕着如何最大可能地保存、恢复或识别语音中蕴含的语义这一中心。而音乐则不同，它往往要求完整保存和逼真重现演奏乐器的音色音调及整首曲子的演奏过程。作为三种声音媒体之一的音频则往往与视频联系在一起，与各种广播联系在一起（如调频调幅广播等）。

最后需要说明的是，尽管这三种媒体分别称为文、图、声，但实际上在计算机内部都表现为简单的数字电信号，都是通过 0 和 1 的编码来表示的（当然可以肯定的是不同媒体的编码方式和处理方式肯定不一样，从而也就可以使计算机能把它们相互区别开来）。只是最后当计算机把它们输出来给人看时，又把它们转化为了人们熟悉的文字、图像和声音。

第四节　多媒体技术的过去、现在和将来

多媒体技术的历史并不长，从出现位图概念处理图形到现在也不过十几年，但与多媒体有关的人机交互技术的发展则稍微长一点。

最初计算机输入输出信息是用穿孔卡的，它上面用打孔来代表二进制数字。自然对机器来说，阅读这种卡没什么困难，但对人来说阅读这种穿孔卡无异于看天书，十分不便。显示器和键盘等工具的出现多少解决了一些人机交互的问题，同时也发明了用文字符号作为信息媒体，这使人们阅读起来相对顺手了一些。

但随着社会的信息化，越来越多的人要使用计算机，比如，政府部门要用它处理日益繁多的公文表格档案，公司要用它处理各种账务发票收据，人们平时也要用它来书写信函、观看节目、处理家务等等。于是在计算机功能增强的同时，使用方法却越来越复杂，没经专业训练的人根本无法使用，自然人们提出能否让计算机用人们平常习惯的处理方法来处理信息，这就是多媒体技术发展的社会背景。需求就是动力，有了这种需求，自然

马上会推动技术的发展。

显示器出现之后，计算机已能处理基本的文字信息了，文字符号的输入一般借助于键盘，输出借助于显示器、打印机，但这样输入信息的速度不够快，计算机输出的信息也不太直观生动。大家知道要用键盘快速输入文字是件不太轻松的事，需经长时间的苦练，况且即便如此，速度还是不够快的；用显示器输出文字，需要用眼睛看，但有时人的眼睛要看其他东西或根本看不见，那怎么办？干吗不用用耳朵呢？对了，于是人们试图用语音合成技术来输出文本信息，使计算机具有“说”的功能，当然现在的多媒体计算机不仅能“说”还能“唱”。

为了提高输入信息的速度，方便人们操作，人们在寻求更快更符合人的习惯的输入方法。计算机语音输入便是一种最直接最方便的方法，因为一般人都有说话能力，而且人们输出信息时用得最多的也是开口说话这种方式；另外人们习惯的方法是用笔书写，现在国内比较普遍的笔式电脑就是用笔输入信息的典范；而对于诸如报纸、文件、表格等的快速输入，则要用到各种图形识别技术，如印刷体汉字识别、表格识别、电子签字系统等。

20 世纪 90 年代以后，多媒体技术如鱼得水，迅猛发展，现在以光盘为代表的多媒体系统已趋成熟，这种系统一般均具有高质量的声像文功能。目前，多媒体发展的一个方向是与家电结合，出现了一系列适合家庭娱乐的多媒体家电，如开始走俏的 VCD 系统就是一例；另一个方向当然仍是以计算机为中心的多媒体电脑。目前的困难在于语音输入输出，图像输入及储存等技术尚未完全商品化或尚有部分技术难题未解决。但相信不久

之后，这些问题很快会解决，现在这方面的初级产品已有面市。

多媒体的发展前景是诱人的，各种多媒体系统正在相继开发，本来认为尚是幻想的多媒体会议系统、视频会议系统正在迅速成为现实，而以可视化技术为基础综合其他媒体技术的虚拟现实技术的发展，将把人们带入一个梦幻时代，那时我们不仅可以在家里遨游远古的恐龙世界，欣赏各种各样奇形怪状的恐龙，而且可以深入原子分子内部漫游光怪陆离的微观世界，以至坐在家中做星际旅行，或许突发异想还可以在银河上筑上一条“行星桥”，使“牛郎织女”从此不再分开，使我们祖先的善良愿望终成现实。

第二章　历史悠久的计算机文本媒体

中华民族源远流长，在漫长的历史长河中，我们勤劳智慧的祖先为我们创造了光辉灿烂的历史，留下了卷帙浩繁的文献典籍，从古朴典雅的《诗经》到气势磅礴的二十四史，从斗智斗勇的《孙子兵法》到行事做人的儒家学说，从治病救人的《黄帝内经》到治国安邦的诸子言论……无不凝聚着中华先辈的聪明才智。特别值得称道的是我们朴实睿智的祖先同时也为我们创造了用以记下这些宝贵财富的文字——中文。正是有了这些表现力异常丰富的语言文字，我们才得以有机会欣赏唐诗宋词元曲，聆听四书五经诸子教诲；即使到了如今的信息时代、多媒体时代，文字仍是最大的信息来源，君不见每天天文数字的报刊杂志，时刻不停的电台广播，琳琅满目的各色广告，哪个离得开语言文字？更不用说我们平时的交谈、写信、通电话了。即使是全球互联网络上的信息目前绝大部分也是以文字的方式传输的（当然以英

文为主）。

但是我们古老的文字也一度使我们处于尴尬境地。因为英文打字机的问世，使西方告别了手写时代，步入了文字处理的机械化时代，尤其是发明电传电字机之后，又使英文步入了电气时代，进入了通讯领域。但我们古老的汉文字却因构字复杂、数量众多，一直造不出使用方便的中文打字机，一再地被拒之以文字处理现代化的大门外，就像近代中国一再地被拒之以现代化强国的大门外一样。

计算机的出现，因其强大的运算功能，使信息的处理进入了自动化时代，也为我们古老文字的再生带来了曙光。今天汉字的计算机编码已经完成，除输入问题尚没很好解决外，其他处理方法已臻完善，汉字信息处理的国家标准亦已经颁布，尤其值得一提的是，汉字的激光照排技术已经走在了世界最前列，使我们终于告别了铅与火的时代，迎来了光与电的世纪，我们古老的文字终于一洗往日的耻辱，恢复了昔日耀眼的青春，显得越发老当益壮了。

本章主要介绍文本媒体在计算机中的表示方法及其输入输出方法和设备。下面首先介绍文本字符的计算机内部表示形式，在计算机科学中这称为字符的内部编码。

第一节　字符编码

正如前面所说的那样，文字既是最古老的信息载体，在计算机发明之前早已经存在几千年了，同时令人惊奇的是，它同时也是最早引入计算机的信息媒体，事实上早期的个人电子计算机上配置的也只有文本。那么文本字符在计算机中是如何表示的呢？难道是把“ABC”或者是“一二三”直接放在里面的吗？

事实并没这么简单，但也没想象的那么难，等大家看了下文之后一定会说：啊，就这么简单呀！是的，知道之后想起来确实很简单。但没想到时又是那么的难，这也正是哥伦布的那个蛋所揭示的真理。

大家知道计算机只能处理二进制数字0和1(因为在物理上只有电压为0和不为0两种简单情况)，但通过许多位0和1的排列组合却可以得到天文数字般多的不同种0、1序列，因为序列数量是以2的指数幂的形式递增的。所以通过0和1的排列组合，就可以得到不同的序列用来表示不同的信息，这就是所谓的编码。在计算机科学中一位0或1称为1比特(bit)，而八位0或1称为一字节(byte)，两个字节简称为一个字(Word)等。根据排列组合的知识，我们马上可以得出，一个字节可以有

2^8 种 0 和 1 的不同排列序列，即 0000 0000，0000 0001，0000 0010，………1111 1110，1111 1111 共 256 种。这 256 种不同的排列就可以用来代表 256 种不同的信息。当然如果用一个字来编码的话即就可以得到更多的排列序列，共计可代表 $2^{16}=64k$ 种不同的信息。顺便说明一下计算机科学中常用计量单位 k，在计算机科学中 1k 不同于物理学中的 1 k，前者 1k 是指 $2^{10}=1024$，而后者 1k＝1000。

这样，现在我们已经知道什么叫编码了，下面就开始介绍文本字符在计算机中的编码形式，先以英文为例，后面再说明中文。

英文共有大小写 52 个字母，再加上数字 10 个，其他还有标点符号、特殊符号等总共 128 个，对此编码即可得到美国标准信息交换用编码（ASCII 码），由于只有 128 个字符（ASCII 码表中每一个符号都称为一个字符），所以用一个字节就足够了，而且最高位可不参加编码，因为 7 比特排列组合共有 $2^7=128$ 种不同序列形式，正好与 128 个字符对应，比如字符空格（空格即一个空位置，在计算机也是一个字符）编码为 0010000，大写字母 A 为 01000001，而数字 0 为 0011000 等等。现在大概大家已经想到了英文文本字符在计算机中是如何表示的了，文本在计算机中就是以 ASCII 码的形式存储在计算机内部的存储器中的。举个例子一串文本字符，比如“Good morning!”在计算机中就是以 ASCII 码的形式逐个顺序放置的，即先放字母“G”的 ASCII 码，再放“O”的 ASCII 码……最后放的是“!”的 ASCII 码，当然为了与其他文本字符串分隔开，“!”后还放了一个作为

结束标志的特殊字符的 ASCII 码。

前面提及 ASCII 码表中仅有 128 个字符，只用了一个字节中的 7 比特，如果再用上这最后 1 比特，就有 256 个字符可表示了，这种扩展了的编码方式得到的字符表就是扩展 ASCII 码表，一共有 256 个字符，扩展的字符包括许多制表符（制表格用），数字符号和希腊字母 α、β 等。

讲到这里，也许大家会奇怪，你说文本字符在计算机中是以 0 和 1 的形式编码存储的，为何我在计算机显示器上或打印纸上看的却是“ABC”而非其 ASCII 码呢？提出这个问题是自然的，具体详尽的回答在本章最后一节中能找到，这里先略微讲一下，以解大家心中疑团。计算机在存储、处理文本信息时确实用的是 ASCII 码，因为这符合计算机内部信息的表达形式，便于计算机处理。但对人而言，阅读 ASCII 码则是件令人苦恼的事，所以计算机在用文本输出信息时，并不直接用 ASCII 码，而是作了一个转换，把它转换成了平常我们能看的懂的文字形式，这一转换就是答案所在。

上面我们已经谈了英文文本在计算机中的表示形式，下面就谈一下我们古老的汉字是如何表示的。

事实上，尽管汉字是方块文字，英文是字母文字，两者相差甚远，但在计算机中却表现为相同的形式——0 和 1。只是英文只有 26 个字母，汉字却成千上万，即使常用的汉字也有六七千，这就决定了两者的内部表现形式并不相同，也正是由于这种不一致，导致了软件使用中经常碰到的中、西文兼容问题，使软件制作者有时不得不发布中、西文两个版本。

这种差异也许聪明的读者已经看出来了，这就是汉字需用两个字节，即一个字的0和1来编码。上文已经提到，一个字的0和1可以有$2^{16}=64k$种不同的编码序列，这已可覆盖平常所碰到的汉字了。不过在实际编码时为方便查找，是把字符按一定的区位排放的，按此方式编码所得的汉字表（也包括英文字母，标点符号等一些常用符号）就是中国国家信息交换用标准代码，简称国标码。另外，因其是用于计算机内部表示文本信息用的，所以有时也俗称内码。自然相对于内码，还有其他汉字编码，如汉字输入用的汉字输入编码方案（如著名的五笔字型、大众码等），还有下面要介绍的汉字字型点阵编码、打印用点阵编码等等。

第二节　文本输入

在上节讲了文本的计算机内部表示之后，读者也许会马上想到这样一个问题：文本是如何输入计算机的呢？本节就是回答这个问题的。

文本输入最常见的工具就是键盘，下面就先简单介绍一下键盘输入字符的原理。键盘的原理不太复杂，大家可以看到键盘上有一个个按键，实际上按键下面还有网格状的电路。当一个按键未按下时，其下网格电路纵横两个方向电流为0；当该键

被按下时，使下面的网格交叉点接通，从而纵横两个方向有电流流过，而键盘自带的微处理器每时每刻都在监测纵横两个方向的电流变化，因而能监测到该键被按下。当然键盘微处理器记录下的不是该键的位置，而是该键的键盘扫描码（一种键盘按键的编码方式，现在我们已经碰到了很多种编码方式，请大家不要混淆）。键盘把扫描码传输给计算机，由计算机的键盘处理程序把扫描码转换为相应的 ASCII 码或执行相应操作（如果该键不是字符键的话），这就完成了一个字符的输入。

中文输入的基本过程与西文相似，但由于汉字在读音外形结构上的复杂性，现在尚没有一种有效的用于键盘快速输入汉字的方法。目前的文字输入方法不是难学就是输入速度慢，比如用拼音输入，倒是十分容易，只要学过拼音并懂得该字的读音就行了，但汉字同音字太多，输入拼音后尚需从众多的同音字中选择出所需的字，速度相当慢；再比如五笔字型方法，平均只需敲四下就可输入一个汉字，但由于其字根众多，拆字复杂，要想熟练地掌握该方法并不是一天两天的事。由于汉字输入方案众多，输入的过程也不尽相同，但基本上可以分为三步。第一步是获得按键的扫描码，第二步把扫描码转换为构字部件的编码（构字部件如拼音法中的拼音字母，五笔字型中的字根），最后根据该字几个部件的编码合成该字获得内码，完成一个汉字的输入。

可以想象得出，当用键盘输入汉字时，由于输入速度取决于人的反应速度，输入速度不可能很高，因此当要输入大量文字信息时，需占用大量的人力和机时，事实上汉字输入确实已成为中文信息处理中的一个瓶颈，但是人们在日常生活中用语言交流

时并没有这种困难啊，这是为什么呢？原来在日常生活中人们用语言交流信息时采用了几种很自然很灵活的方法，这就是我们熟悉的用嘴说，用手写或者印在纸上。于是人们在想：能不能让计算机也能接受我们这种信息的交换方式呢？这样的话，我们输入信息不是又快又省劲，且自然多了吗？需求是科学发明的动力，有了需求，人们自然会去努力探索。经过几十年的风风雨雨，人们终于获得了初步成功，使计算机也开始"牙牙学语"，开始学会"听"人说的话、"看"人写的字，这就是计算机口授输入、笔输入、印刷体文字识别等技术。

口授输入技术采用的是语音识别技术，通过专门的语音识别装置，把人说的话直接转换为文本存入计算机，有关语音识别技术的介绍请读者参见声音媒体一章语音识别技术部分。笔输入和印刷体文字识别都采用了图像文字识别技术，但两者的具体方法和技术难度并不相同。下面以汉字识别为例介绍这两种方法的基本原理和过程。

汉字识别的基本原理实际上讲起来并不复杂，就是一个汉字特征提取和匹配的过程，这正像我们用字典查一个生字一样，只是我们这种过程是潜意识进行的，我们并没有感到什么特征提取、特征匹配的过程，但让计算机识别时，这就变成"有意识"的了，因为我们得告诉它如何去做。

所谓一个汉字的特征就是一个汉字能与其他汉字相区别开来的特征，比如其笔画和笔顺，其四周特点、四角特点等。打个比方例如"日"和"曰"两个字的基本区别就在于"日"字较"瘦长"，而"曰"则较"矮胖"，当然这种描述计算机是看不懂的，需进

行数字化翻译成计算机能看的懂的 0 和 1。汉字匹配最简单的方法就是根据提取的汉字特征值去“查字典”，找出汉字特征库中存储的与该字特征最相近的一个字来，这个字就是识别出的结果。目前汉字识别的方法很多，下面介绍一下笔输入和印刷体汉字识别技术。

在笔输入系统中，当特殊的书写笔在输入板上书写时，计算机根据书写板上留下的轨迹，识别出汉字的笔画和笔顺，以此作为汉字的特征，然后与汉字特征库中存储的数字进行匹配就可以了。这种联机笔输入系统技术上相对容易，目前已有实用产品面市。而印刷体汉字识别由于不可能获得及时的书写信息，识别就比较难了。它的输入设备一般是扫描仪。扫描仪把一页印刷体文稿扫描入计算机后，计算机首先分割出单个汉字，然后对每一个字进行识别，其汉字特征一般取汉字的外围特征或四边特征，如果能设法利用上下文信息自然效果更好。提取出特征后接下来就是匹配找出最相似的字来。

汉字识别（包括其他文字识别）说起来很简单，但实现起来并不容易。首先以什么作为特征，如何提取特征就是一个难题，大家知道可以提取的特征很多，但要提取出能把所有汉字两两区别开的特征几乎不可能；另外如何在庞大的汉字特征库中快速找出最佳匹配的汉字来也是一个很值得研究的问题，否则识别速度会很慢。

但汉字识别的好处是显然的，代表着汉字输入的一个方向。所以近几年来国内外投入了大量的人力物力研究这个问题，并取得了一定成就。可以期望，不久之后更实用的系统将会面世，

那时我们古老的文字将更加容光焕发，老当益壮！

第三节　文本输出

谈了计算机文本输入的各种方式之后，自然又提出了文本输出的问题，即计算机是如何把信息用文本的形式输出给外界的。这个问题在本章第一节中已提及，这儿就详细讨论一下。

目前文本信息的基本输出方式有三种：显示器输出、打印输出和语音输出。语音输出就是计算机模拟人说话的方式用有声语言输出文本信息，这实际上已牵涉到了声音媒体中的语音合成技术，这儿暂不细谈，把它放在声音媒体一章中再细述。下文主要介绍文本显示的原理，顺便谈一下打印输出的基本方法。

要谈文本的显示原理，我们预先不得不了解一点有关光栅显示器的显示原理，因为光栅显示器是目前个人计算机上主要配置的显示器。

光栅显示器的显示原理与电视机显像管相似，也是通过电子束撞击荧光屏上的荧光粉发光的，只是显示器显示的不是电视信号，而是储存在显示器存储器中的文本和图像信息。由于显示器是用电子束撞击发光的，大家可以想象得出，显示器上显示的图像肯定不是连续的，比如一根水平直线，就是由许多极小的光点依次水平排列形成的，只是由于光点太小我们的眼睛不

能分辨开来，才使这串水平光点看上去像一条直线。在计算机图形学中，这种小光点称为像素，显示器上的一个点就是一个像素。我们可以直观地想象如果一幅图像用较多的像素来显示，那么它可能会清晰一些。事实正是如此，在计算机图形学中显示器一屏像素的多少可以用分辨率来描述。比如分辨率为 320×200 就表示该显示器在这种分辨率下水平一行可以显示 320 个像素，垂直一列可以显示 200 个像素，整屏共有 320×200=64000 个像素。320×200 的分辨率是比较低的，目前好的彩色显示器分辨率可以轻松地达到 640×480，甚至 1024×768 等。另外一个显示器的分辨率、颜色数等是可以自由设置的，每一种可选的设置称为一种显示模式。

当计算机显示器用于显示文本字符时，实际上显示的是字符的“图形”，比如显示字符“F”时，就在显示器屏幕上画一个“F”的图形。只是由于字符形状是一定的，可以预先设计好放在计算机存储器中，当要显示时，只要按字符索引找到内存中这个字符的点阵信息放入显示器存储器中就可以了，点阵信息的显示（即相应像素的点亮与否）则由显示卡自动完成。字符点阵的大小一般是固定，常用的是 16×8 点阵，就是横向占用 8 个像素，纵向占用 16 个像素这样一个矩形区域的大小，当然也可以有其他大小的点阵。

我们用涂黑的小黑格代表显示字母“F”时应点亮的像素，由此组成字母“F”的图形。计算机在储存字符点阵信息时，也要进行编码，把字符的点阵信息转化为 0、1 序列。计算机编码时常用一个比特位对应一个像素，若对应某个像素的比特位为

1 则认为显示该字符时该点像素应点亮，为 0 则不用点亮，至于具体显示时该用什么颜色显示则由用户编程决定，我们这里就不谈了。按此编码原理，字符“F”的点阵信息可以编码为（按从上到下依次编码）。点阵第 0 行编码为 01111111（从第 7 位到第 0 位排列），第 1、2、3、5、6、7、8、9、10 行点阵码值相同，均为 00000001，而第 4 行点阵码值为 00111111，第 11 行点阵码值为 00000000，共需 12 个字节。类似对于 16×8 点阵编码共需 16 字节。由于西文字符较少，所以其常用的点阵信息一般是做成硬件直接存在计算机只读存储器中的，而汉字因数量众多，一般常用的也六七千之多，所以直接存于内部存储器中很费存储空间，因而一般是放在专门制作的汉卡或外部存储器如硬盘、软盘上的。自然字符点阵信息的存放也是很讲究的，需按某种规律存放，这样有利于快速查找，比如西文可按“ASCII 码大小先后存放，而中文则可按区位码放置，因为汉字内码与区位码有某种确定的对应关系，而汉字在计算机中正是以内码表示的。

文本字符的显示装置除光栅显示器外还有很多，比如便携式计算机、笔记本计算机中常用的液晶显示器，大屏幕广告牌上常用的发光二极管显示器等，这里不再赘述，下面再简要介绍一下文本打印输出的方法。实际上文本打印输出的也是文本字符的点阵信息，只是这种点阵信息是专为打印机编制的，适合于打印机处理。说到这里，大家也许就明白了打印字符的方法，即如果代表点阵信息的某比特位为 1，打印机的打印器就打印该位，否则就不打印。当然打印机打印时往往不是一点一点打的，否则速度太慢了，而是一排比如 9 点或者 24 点一起打印的，这

在针式打印机中就称为9针或者24针式打印机。喷墨打印机与针式打印机类似，只是它是用细小的喷嘴喷墨水，而不是像针式打印机那样用钢针击打色带，激光打印机的打印方法与上述两种不太一样，它有点像静电复印，所以比起前面两种打印方法来速度要快得多，而且打印质量也高。另外还有热腊打印等，这里不再详述。

另外，字符显示或打印时除用点阵信息外，还可以利用矢量字符信息，矢量字符也是一种描述字符形状的方法，所谓矢量就是既有大小又有方向的量。用矢量来刻画字符形状有点像汉字需用笔画和笔顺来描述一样。举个最简单的例子，汉字“一”只有一个笔画，就可用由该笔画的起点和终点位置，笔画的方向等构成一个矢量来描述，其他笔画较多的字符的描述方法也一样，只是这样的矢量要多一点。事实上，对于矢量字符的描述有一套专门的描述语言。矢量字符比起点阵字符来有很多优点，首先前者可以放大缩小而外形基本保持不变，而点阵字符一旦放大过多，则会出现锯齿形，很是难看；另外矢量字符可以把字符外形做得很好看。所以矢量字符多用于美术设计及大号字的显示和打印上。

最后提一下，对字符外形更细致的描述可用曲线来拟合，这里不再细述，但在实际的多媒体电脑中，这种字符显示和打印已用的很普遍了。

第三章　形形色色、亦真亦幻的计算机图像媒体

如果大家看过科幻片“面具”的话，一定会惊诧于片中面具人变幻莫测的脸部表演：他一会儿神气地把鼻子拉得长长的，一会儿又吓唬人似地张开血盆大口，一会儿又生气地把眼珠子伸出眼眶之外犹如螃蟹的眼睛……你一定会怀疑这演员的脸真经得起这般折腾吗。再比如在“小精灵”一片中，小精灵和他的叔叔们白色雾状的躯体以及可爱的小精灵乌黑漂亮的大眼睛害羞的样子一定给你留下颇深的印象，于是你又疑窦丛生，难道拍电影时他们真请来了精灵帮忙？如果你还看过“侏㑩纪公园”的话，你一定更加惊奇了：恐龙不是早灭绝了吗？难道拍电影的又打外星球运来了恐龙不成？而且又怎么让这些巨型怪兽乖乖地在摄影机前表演呢？

如果想要回答这些问题，等你看完本章的内容之后就不难了，因为它们都是综合使用了计算机图形、图像、动画、视频等技

术之后创造出来的，而介绍这四种计算机多媒体技术正是本章的任务所在，而且在本书中这四种多媒体被统称为图像媒体。当然图像媒体中的“图像”要比这四种技术之一的“图像”外延要广一些，注意区别。

第一节　乐于助人的计算机图形

也许有些读者在计算机上玩过简单的海龟作图，也许有些也玩过复杂的三维动画，当然更多的读者也许都玩过精彩的电子游戏，游戏中那离奇的故事、精彩的画面、非凡的动作，常引得同学们废寝忘食，耽误了学习（这不太好），所有这些从最简单的画线画图到最复杂的三维动画，都涉及到了计算机如何作图的问题。本节就谈谈这个问题，它也是计算机图形学的主要内容。

一、计算机是如何生成图形的

计算机显示一个图形的主要设备有显示器、绘图仪、打印机等，最常用的就是显示器，显示器的种类很多，比如常见的光栅扫描显示器、液晶显示器，其他还有刷新式矢量显示器、存贮管显示器、气体放电平板显示器、发光二极管显示器等，各应用于

不同的场合。下面主要介绍一下个人计算机、工作站上用得较多的光栅扫描显示器上生成图形的原理。

在前面一章中大家已经简单了解了一些计算机在光栅扫描显示器上显示一个图形的原理，知道了一个图形是由众多的不同色彩的像素构成的，也即像素是计算机能显示的最基本的图形，因此显示或不显示一个像素点就成了计算机用光栅显示器显示图形时最基本的操作，因为任何复杂图形都是由许许多多的像素组成的，都可用这种基本操作的有机组合来实现（如何有机地把这种基本操作组合起来构成欲显示的图形，就是图形的显示算法，这也是计算机图形学的主要研究内容）。比如如果知道了一条直线的起点位置和终点位置的坐标及欲显示的颜色，那么简单地使用一条循环语句，依次点亮与起点和终点在同一直线上的各个像素，显示器就显示出了一条直线的图形。这里提到了点的坐标，就顺便说明一下图形显示的坐标系统。图形的坐标系统一般有两种，一种就是我们通常所使用的坐标系统，它可以根据我们的需要任意调整R度及原点位置，这可称为客户坐标系；另一种就是显示器的屏幕坐标系统，由于分辨率的关系，它的数轴的长度，原点位置一般是固定的。计算机作图形时，就是要把我们头脑中的客户坐标系中的图形（二维的或三维的）映射到屏幕坐标系（二维的）上。屏幕生标系的原点一般取在左上角，并以水平方向为X轴，垂直方向为Y轴，一个像素位置就代表一个坐标点，比如分辨率为640×480的显示器屏幕，X方向的坐标值就为0－639，Y方向的坐标值为0－479，而且都取整数值。

上面已经说了，计算机显示或者说生成一个图形的最基本的操作是显示一个像素点，这一般由硬件来实现，所以显示一个图形时人们要做的就是图形的显示算法了，即如何用这种像素点的基本操作的有机组合来显示一个图形。

计算机图形的显示算法是计算机画好一个图形的关键所在，对于规则的几何形体可借助于其数学方程来推导出其显示算法，所以不同的形体往往具有不同的算法，即使同一形体也往往由于使用的数学模型的不同而不同，这正如在解析几何中直线可以用点斜式、两点式等不同的方程来描述一样。与之不同的是，计算机作图时不同的算法画出的图形往往效果很不一样。比如画两条斜线，在这张示意性的图中可以看出，由于内插过渡点的算法不同，导致两条直线看上去效果不一样，一条看上去更像是一直线，而另一条直线则有较明显的锯齿形出现。

上面介绍的是规则的几何图形的显示算法，对于不规则的复杂曲线和曲面，则要用曲线拟合和曲面拼接的方法。曲线拟合就是用几段曲线的光滑连接来逼近复杂的不规则曲线；曲面拼接的原理也差不多，只是它是用几个小曲面的拼接来构造复杂曲面而已。以二维图形的曲线拟合为例，比如一片叶子，他的画法有些类似于画家画素描，先用几段折线搭出它的大致形状，再用内接于折线的几段光滑曲线的拼接来构成整个复杂图形。

从上面介绍的几种图形画法可以看出，要想在计算机上画好一个图形，并不容易，需要作者综合考虑多方因素，比如分辨率、彩色数等显示器特性，还有图形的几何特点，计算机的性能等因素，然后才能提出一个合适的算法，而且这往往还是一个反

复的过程。所以当你坐在计算机前轻松地欣赏屏幕上那一幅幅奇妙的画面时，不知想到过没有，这其中凝聚着作者多少心血，耗费了作者多少时间。

上面花了不少篇幅介绍计算机作图的基本原理，其复杂的过程和需要考虑的众因素，是不是令你有种可望而不可及的感觉？确实要制作出一幅精美的图画并不太容易，要花费很多时间和精力，并还要你懂点计算机科学、计算机图形学、色彩学、数学等学科的基本原理，而且还要求你有一定的美术知识，够“交叉”的了。好在由于技术的发展，使人们能开发出各种专用的和通用的图形制作系统，让这众多的因素和算法交给计算机去处理，留给人们要做的只是创造性的劳动，即使没有太多的美术知识的你，也可以充分发挥你的想象力，利用图形系统提供的作画工具在屏幕上尽情挥洒。画得不好也不要紧，只要你有足够的耐心去修改就行了，而且这种修改往往是轻而易举的，按几下鼠标就可以进行擦除、移线、上色、转过一个角度、推远、拉近等等操作，并且可以反复无数次，而根本不用担心会弄坏“画纸”或无法再上色了。计算机最后还可以帮你进行消隐、明暗、纹理等处理，使你的画面形象看上去更逼真更具“专业”水平。于是当一架架逼真的 F117 隐形飞机，当一个个我们想象中的宇宙勇士跃然于屏幕上时，作图时的繁琐、紧张、失意又算得了什么呢，留下的只是浪漫、惊奇和成功的喜悦！

二、如何使计算机生成的图形更逼真

不知大家注意到没有，如果老师让我们画一个橘子的话，面对眼前的橘子，我们除了把它画成圆形之外，还要尽力去表现橘子粗糙的表皮，而且还要分析桔子表面各个部分的明暗关系、色彩的浓淡，最后还不忘在橘子背光面之后添个阴影，这是为什么呢？事实上我们这样做无非是想把橘子画得看上去更像一个橘子，而不是苹果或者梨，即真实地把它表现出来，使画出来的东西看上去显得逼真。

在用计算机画图时同样也有这样的要求，而且其要求更高，比如由于它是画在屏幕上的，我们无法观察其被挡住的部分，于是我们还要求画面上的东西能按我们的要求任意地转过来转过去，使我们能观察其被挡住的部分，这在传统画面上是不可能做到的。这部分涉及的技术主要有图形变换、消隐、明暗色调处理、阴影生成和纹理处理等。

所谓图形的变换就是将已画好的图形经过放大、缩小、旋转等操作，插入到新图中，以提高制图速度。由此可见，这里的图形变换实际上可为我们提供更多的功能，而不仅仅限于用来观看画面上物体被挡住的部分。比如一张很大的电路原理图（或旅游图），如果整张图显示在一个屏幕上，虽可窥其全貌却难以分清其细部，这时如果运用画面的缩放功能，就可对其某个局部

进行放大操作，于是我们就可看清每一根连线每一个元器件了。另外变换还可以用于产生各种动画效果，比如一个简单的例子是雪橇在雪面上水平滑过去的动画画面，这只要简单地多次反复利用平移变换就可实现了。

图形的变换很多，有简单也有复杂，但由于一个复杂变换常可看作多个简单变换的叠加，所以我们有可能把千变万化的变换归结为几种最基本的变换，对于二维图形其基本变换有平移、绕原点旋转、对称变换、变比等，而对于三维图形则还有绕坐标轴旋转、透视变换等。对于上述大部分基本变换大家可以从其字面上看出具体变换形式，需要说明是变比变换和透视变换。变比变换是指对图形的各个部分按不同的比例进行放大或缩小使之形状发生畸变的变换，而透视变换，则是把图形按透视关系变换，使之符合透视原理。下面再举两个由基本变换构成复杂变换的简单例子。比如有时我们希望图形绕某个给定点(m,n)旋转θ角，而不是绕原点，那我们怎么办呢？我们可以这样做，先把物体平移至原点，然后把平衡后的物体绕原点旋转θ角，最后再把旋转了θ角的图形平移回(m,n)处，就实现了绕该点旋转θ角的目的。再说我们一开头提出的问题，即我们想观察某个图形背面的一个小凹痕，比如画面上月球背着太阳一面的一个小环形山，这只要先把该图形绕某个轴旋转一个角度让凹痕转到不被挡住的地方，然后再施行变比变换，把小凹痕放大以便于我们观测，由上两例可见由几个基本变换确实可以复合出一个复杂变换。

那么计算机对于各种变换是如何处理的呢？即变换在算法

上如何表示呢？实际上所谓图形变换无非是把图形上的每个点都做了该种变换，而坐标点的变换在数学上表现为矩阵运算，所以计算机实现图形变换时在其算法上就表现为大量点的矩阵运算。由于矩阵运算运算量是很大的，所以可以想象得出，对于表面极不规则的复杂图形由于无相关性可寻必然导致数量众多的坐标点的矩阵运算，运算量是相当大的，一般的个人计算机往往难以胜任，所以对于复杂的图形处理常要用到图形处理速度比个人计算机快得多的图形工作站。

为了使计算机画出的图形更逼真，我们前面已经说了，尚需对图形根据实际的光照和材料的质地做出各种明暗和纹理处理，下面就谈谈这方面的图形处理技术。

首先为了使画出的三维图形具有立体感除了应用透视变换外，还要把图形背面、侧面被挡住的线和面裁剪掉，因为物体一旦被挡住就看不见了，不应再显示出来，这种图形处理技术常称为消隐，即消去隐藏面的意思。另外还要处理好图形表面的明暗关系和色彩的浓淡关系，比如光源直接照到的地方往往比较明亮，照不到的地方比较暗淡，而且物体背后光照不到的地方还会形成阴影，这些现象在计算机作图时都应表现出来。最后就是要使屏幕上的物体看上去具有质感，即充分反映物体表面的材料质地，比如橘子皮的粗糙感，苹果表面的光滑感等。图形只有进过这些处理之后，看上去才会显的逼真，生动。当然这些处理又都涉及到复杂的算法，不过现在我们作图时基本上不用去管它们了，因为如今的图形系统基本上都提供了现成的算法供您使用，您只要做简单地选择之后，设置一下参数就行了，这在

图形生成部分我们已经都提到了。

三、计算机图形能为我们做什么

计算机作图到底有什么用呢？当然不会仅仅是好玩，要是这样，也太小看它了。事实上计算机图形可以为我做很多事，比如计算机辅助教学，计算机辅助设计等计算机辅助领域，大家也许还记得当我们刚开始接触立体几何时，常常为想象不出三维几何图形的形状而烦恼，有了计算机图形你的这种烦恼就可一扫而光了，因为对于显示简单的三维形体对它来说不过是“小菜一碟”，实际上在许多数学软件中几乎你一给出函数或方程，它就能马上为你画出对应的画形来。另外应用于教学方面的计算机图形还能为我们生成多种挂图，制造多种幻灯，甚至能模拟实际过程让你在计算机屏幕上去解剖一只兔子、一只青蛙，抑或是我们平时连靠近它一点都不敢的鳄鱼、老虎。如果你想的话，甚至还可以“乘船”畅游人体的血液循环系统呢，当然也是计算机模拟出来的，想想看，这该有多么奇妙啊？而且计算机图形不仅可以应用于教学，还可以应用于生产领域，提高生产率，创造经济价值呢。比如在电子线路设计自动化中，你可以直接在计算机上画原理图，而修改是十分方便的；画完之后，计算机可自动进行电气校对，自动布线甚至最后直接生成硬件电路。这样，过去需要花几个月甚至几年才能完成的设计上工作今天借助于这种计算机辅助设计工具几天就可以完成了，而且质量更好，出错

的可能性更小，这在以前是连想也不敢想的呀！科学技术可真是第一生产力啊！

计算机图形除了在计算机辅助领域发挥重大作用外，近年来又在计算机动画、电子游戏、电子广告、多媒体演示、虚拟现实环境等方面获得了巨大应用。随着今后图形技术的发展，计算机图形将在越来越广泛的领域获得应用。

第二节　神通广大的计算机图像

也许读者曾不只一次地在电影电视上看到过，喜欢侦探小说的读者更应注意到，在发生凶杀案之后，警察往往要封锁案发地点并进行严密保护，因为其中也许就隐藏着罪犯的蛛丝马迹，而其中最常见的往往是罪犯留下的脚印和指纹。对于指纹，读者绝不会陌生，随便伸出一个指头，在一张干净的白纸上按下，便留下了你那独一无二的螺旋形图案——这就是你的指纹，而且这仅属于你。由于每个人的指纹不相同，所以指纹便成了可以证明本人身份的最特殊也是最一般的标记。这不。在“白毛女”中，黄世仁就强迫杨白劳在喜儿的卖身契上按红色的指纹作为签订契约的证明。在旧社会，指纹成了“黄世仁”们欺压贫苦大众的帮凶，而在今天，指纹则成了公安干警追查罪犯的有力证据。所以在案发现场，常可发现技术人员拿着放大镜在仔细观

察门窗、桌面，甚至罪犯有可能碰过的酒杯、茶缸，扔下的纸片香烟头，其目的就是想发现案犯留下的指纹等蛛丝马迹。

但同时细心的读者也会发现，过去在发现罪犯指纹之后，并不能马上确定谁是它的真正主人，而只能在确定的嫌疑犯中慢慢核对，如果一时没发现嫌疑犯，核对也就无从谈起，出现这个问题的关键就是在于当时都是用人工的方法鉴别指纹的，不仅技术要求高，而且工作量大，所以即使知道了指纹，也不能很快鉴别出它的主人。

但现在不一样了，应用图像技术，在登记户口建立每个人的户籍档案的时候，如果保存一份该人的指纹并把它随同其他户口信息一起存入计算机，那么当要确定一个指纹属于谁的时候，应用计算图像识别技术，让运算速度极快的计算机去核对比较的话，那识别速度就快得多了，更厉害的是计算机根本不会疲劳，可以夜以继日地工作。所以现在要鉴别一个指纹已不是什么难事了。当然如果要查找的人群范围越大，则可能花的时间就越多。具体鉴别时，可以先在有前科的人群中查，或嫌疑犯中查，以缩小查找范围。据说美国一个州就用这种指纹鉴别技术成功地确定了30年前一桩凶杀案的杀手，只是罪犯被抓获时早已成了一个不中用的老头了，如果此案发生在现在，那他就无法逍遥法外这么多年了。

您瞧，计算机图像识别技术是多么的厉害。当然计算机图像(这里指静止图像)的用处远不只指纹鉴别，还可以用以鉴别签名，识别表格，使大量的文件处理实现自动化。另外现在以光盘为存储介质的“电子相册”也已问世，用这种相册，你就可以在

计算机显示器上一张一张地翻阅自己往昔的身影，回忆那一幕一幕有趣的往事和美好的日子。当然电子烹调书更可为你的每日三餐增色不少，CT 造影又为你解除疾病之苦提供了方便，还有电子导游用，电子地图，电子生活指南丛书……无不为你的工作、生活、旅游提供了极大的方便，当然这其中都包含着计算机图像的默默奉献。

本章这部分就让我们来谈谈计算机图像（静止）处理的一些基本概念和方法，使我们对这多姿多彩的计算机图像有些深入的了解。

一、计算机图像的来源

计算机图像既然能干这么多有用的工作，那么先让我们来看看它的图像是如何输入计算机的。

在计算机科学中广义的图像包括本章所有的媒体内容，即指包括图形、图像（静止）、动画和视频图像的图像媒体，不过这部分讲的计算机图像主要是指静止图像，即画面上的物体都是静止的，不活动的，下面用到图像这个术语时就不再指出这一点了。

静止图像严格说也就包括图形、文字，它们已在前面讲过了，这部分就不讲了，所以这部分讲的图像一般指的是扫描仪等图像输入设备、专门的图像记录设备如造影系统，遥感设备等扫描或记录下后输入的计算机的图像，以及从活动图像上截取下

来的静止画面等。

扫描仪是一种通过光传感器把光信号轻变化电信号的图像输入设备，与摄像机不同，它只能处理静止画面，如相片、图片、报纸等。所以扫描仪一般多用于照片、图片的输入，在印刷体汉字识别，签名识别，表格自动输入等图像识别领域一般也用它作为输入设备。扫描仪的最主要指标是其分辨率和彩色数(单色图像用灰度级表示，即有多少级灰度)。与显示器不同，扫描仪的分辨率用 dpi(dots per inch 的缩写，即每英寸的点数)来衡量，比如 300dpi，就是说每英寸可以分辨出 300 个点。彩色数是扫描仪能区分的最多的颜色数，一般用需要几位比特位来编码代表，比如真彩色需要 24bit 来编码，即其可区分出 2^{24} 种彩色。扫描仪一般有台式和手持式两种，台式一般分辨率比较高。

专门的图像记录系统顾名思义一般是应用于专门领域的，如 X 光断层扫描成像系统可用于疾病诊断和材料探伤，红外遥感用于记录地形地貌、气象农业资料等等，这里不再一一列举，视频截取一般需要用到视频捕捉设备，这放到视频部分再讲。

二、图像是如何储存的

当我们在计算机上一张一张翻看电子相册时，当我们在计算机上浏览世界风光时不知大家想过没有，他们是如何保存的呢?

事实上图像的储存一直是个大问题，问题的本质在于图像

的数据量太大，举个简单的例子，对于一张 8.5×11 英寸的彩色照片，如果用 300dpi、24bit 真彩色的扫描仪扫描输入的话，那么这张相片需要用 8.5×300×300×24bit，大约等于 25MB(1MB≈10^6 bytes≈10^3KB，B 代表 byte，即字节，KB 为千字节下面还用到的 GB，1GB≈10^9bytes≈10^6KB≈10^3MB)。也许大家对此没什么概念，但只要想想一台配置已算不错的 PC 机一般内存最多也只有 16MB，连一张相片的数据都存不下，那么可以想象出图像的数据量也实在太大了，即便是海量存储器 1GB 的光盘，最多也只能有储 40 张这样的照片。如果我们要把这张相片传送给远方的亲人，用现在比较常见的调制解调器速度 9600bps(bps 是数码率的单位，1bps 意思是每秒传送 1 个比特位，即 bits per second)来传输的话，则需要花近 6 个小时，即使在速度比较快的局部网上以 10mbps 的速度传输的话，也需近 20 秒钟，这样长的时间当然是难以容忍的。

也许你要问了，那么电子相册上这么多的画面到底是如何储存的呢？办法只有一个，那就是设法减少图像的编码数据量，即必须采用压缩编码技术。虽然图像数据看上去杂乱无章，但是从计算机的内部表示看，我们前面说了，都是表现为有限位的 0 和 1，按照统计学的观点看，必然存在一些像素或图像的某一小块区域是相似的，比如红色画面上，必然有很多块是红色，蓝色画面上，必然有很多块是蓝色的，另外我们人的视觉和分辨率是有限的，有时图像数据去掉一些，在计算机看来图像已经有些改变了，而我们人眼却根本感觉不出变化。因此，可在允许一定的失真的条件下，对图像进行压缩编码，以大大节省储存空间。

图像的压缩编码有多种方法，如利用相邻像素的相关性进行压缩，进行各种变换后抽取出特征值进行压缩，采用矢量量化进行压缩等等。目前静止图像的压缩已有国际标准，比较著名的就是 JPEG 标准，压缩比可以达到 15 : 1 甚至 25 : 1。如果用 25 : 1 的压缩比压缩前面那张相片的话，压缩后数据量就只剩下 1MB 了，1GB 的光盘就可以储存 1000 张这样的相片，效果是非常明显的。当然即使压缩后图像的数据量仍是比较大的，用一般的软磁盘存储还是比较困难的，因为一张 3.5 寸的软盘充其量敢只有 1.44MB，只能存上面那样的一张相片，事实上在多媒体系统中，用的最多的就是大容量的光盘，一般其容量都在 600MB 以上，基本可满足多媒体存储的需要，缺点是目前的光盘一般是只读的，即只能读出数据，而不能写进去数据。有关光盘存储技术我们专门在第五章用了一节篇幅介绍，请大家参考。

三、计算机图像的处理

对于已存储在计算机中的图像的处理一般包括三种，这就是用图像的编辑、变换及理解(或称识别)。

对于静态图像的编辑处理一般包括修改、裁剪、添加、组合、拼接等。这通常是在图像编辑软件的帮助下进行的。以设计一本书的封面为例，手头有好几张画准备在设计时用，但又不能全用，于是我们组合了一下，打算设计这样一种效果，最前面的一张画全部显示，其后面的画面依次露出一部分以表现被前面的

画面挡住了。当然原先的画面很大，我先把它们依次缩小，把最前面的画直接放上去，而后面几张需先经裁剪，剪去不露出来的部分，然后添加到封面上。当然画片不可能一下子放好，我就需要不停地修改直至满意为止。最后还要添加上书名、著者等文字信息。这就是综合利用多种图形编辑功能的一个例子。

实际上图像的变换功能常常是与编辑功能揉和在一起的，比如在上面例子中缩小一个画面，另外也许还要把后面的画面转过一个角度斜着放，这都涉及到图像的变换操作，当然这还只是简单的变换，其他变换形式还有修改颜色，修改亮度，图像显示，彩色图像与灰度图像的相互转化，轮廓提取，边界增强，乃至受损图像的复原等。举个简单的例子比如我们想把一幅彩色图像打印出来，但没有相应的彩色打印机，于是我们可以把彩色图像转化为灰度图像，用具有灰度打印功能的打印机打印；但如果我们只有针式打印机，它又不具有灰度功能，于是我们又只好把灰度图像的轮廓提取出来，用针式打印机打印其线框图。这样我们虽然没有把原图像打印出来，但借助图像变换技术把它的轮廓打印出来了，这在某些场合也许就足够了。当然跟图形变换一样，图像的变换也涉及复杂的算法，花费的时间也比较多，但好在现在图像处理工具都提供了某些快速的变换算法，我们用起来还是比较方便的。

图像处理中最复杂的莫过于图像理解了，它要求在向计算机输入一张图片之后，计算机据此能识别出画面上的一个或多个景物，这有些类似于人的视觉功能，事实上也真是如此，因为图像识别正是计算机视觉的基础。前面提到的计算机自动指纹

鉴别就是这样一种图像识别过程，它要求从一大堆人的指纹中，挑出一个与待鉴别的指纹最相似的指纹来，从而根据储存的有关这个指纹的人的信息就可以迅速地找到这个人。这样看来文字识别实际上也是某种形式的图像识别了，确实对于印刷体文字识别，其原理与图像识别是很相似的，可以看成是后者的具体应用。而事实上图像识别的应用是很广泛的，并不仅限于文字识别，例如可用于染色体的分类、排列，血球的分类、计数，航空照片的地貌分析，计算机视觉及机器人识别系统等，但这些应用系统已广泛涉及到具体应用领域，在技术理论上已交叉了多门学科，而不仅仅限于图像处理。比如计算机视觉从理论上就涉及到了人工智能、计算机科学、计算机图像处理、模式识别以及神经生物学、心理学、物理学等，有关这方面的技术已超出本书范围，就不再深入讲了，下面只就一般的图像理解系统的基本原理作一简单介绍。

在用图像理解系统识别图像之时，第一个碰到的问题是我们要识别的目标是什么，是航空照片上的工业区呢，还是军事基地？或者是森林呢，还是农田？在汉字识别系统中就是要首先区分开一个一个的汉字，而在图像识别中第一步就是分割图像，把它分割成一块一块有意义的区域，然后针对这些有意义的区域再实施识别。图像实施分割的另一个目的是把前景景物从背景中区分开来。图像分割的方法很多，比如灰度阈值分割法、样板匹配法、区域生长法等等。灰度阈值分割法原理较简单，适用于前后景灰度在统计上相差比较大的图像，具体方法就是先设定一个阈值，如果背景大部分比较暗的话，那么就认为小于阈值

的像素是背景，而大于阈值的像素属于前景景物，从而经过变换综合可以得到前景景物的有效区域。

区分出有效区域之后，接下来是从有效区域中提取出能比较完备地描述景物的特征值，这一步很重要，往往关系到最后能否准确识别，所以特征不能随便取。最好是该物区别于他物的本质特征，当然这是比较难的。

最后一步跟汉字识别一样，是一个特征匹配的过程，就是用提取出的特征信息与景物库中储存的各种景物的特征值相匹配，找出一个匹配得最好的景物（按某种准则），它就是该景物的识别结果。图像识别说起来原理也很简单，但实际实现时还有很多工作要做，比如如何提取特征，如何建立景物库，如何快速匹配等全都是关系到系统成败的关键所在。

四、结语

本节简要介绍了与计算机静止图像处理有关的各种技术，涉及到了图像获取，图像压缩编码和储存，图像的多种计算机处理和应用等。图像处理还是个年轻的领域，但已显示出它在生产生活中的巨大作用。比如具有视觉功能的机械手已在工厂流水线上大显身手，成倍地提高着生产力；各种电子画册更是包罗万象：邮票、时装、菜谱、地图、植物、动物等等极大地丰富了人们的生活。据报载最近市场上又推出了一种影象杯，它通过计算机和摄像机把买者的肖像记录下来，经计算机处理之后生成胶

片输出。胶片经过特殊处理之后就能把买者的肖像印在了他买下的搪瓷杯上。此后每当你端起这只印有自己肖像的杯子时，又将为平凡的日常生活增添一道幽默的风景线。

随着技术的发展，多媒体将越来越广泛地应用于社会的各个领域，计算机图像作为多媒体技术的一个重要分支将会发挥越来越大的作用，而有志于此的朋友将会在这片领域中取得丰硕的成果，因为这里有一片广阔的天地任你去自由翱翔！

第三节　寓教于乐的计算机动画

初学地理的朋友也许会深有感触，要记忆那众多的山名、地名简直令人束手无策，单全世界主要的 200 多个国家及首都的名称就够你记的了，更不用说那不计其数的城市、地区及其地理气候特征了。记忆可真是件辛苦的事，尤其是死记硬背，更是事倍功半。但如果用游戏的方式来记忆，那朋友们肯定乐此不疲了，效果自然也就不一样了，这是因为人相对更擅长于联想记忆，对于生动直观的东西，往往更容易记住。于是我们可以在计算机上设计一种交互式的动画片，玩这种游戏时，主人公不停地周游世界，每到一个国家，计算机不仅显示出该国的地理位置和名称，还可以介绍气候特征，风土人情（当然这都是用直观生动的图文来显示），甚至还可以在该国各地"转悠"上半天，用这种

方式不仅生动形象，而且还充分调动了人的积极性，不仅可以耳闻目睹，还可以动手操作，效果当然就不一样了。

其实，动画不仅可用于制作这种教学游戏，而且还可以用于演示各种有趣的物理、化学、生物等过程，比如演示植物光合作用的过程，植物根与土壤交换水和矿物质的过程，这都可以利用动画来演示整个活动过程，效果当然也是传统的挂图所无法比拟的。另外有些现象和过程我们是无法用肉眼观察到的，比如原子、分子及内部的微观世界，恒星、星系等宇观世界，为了演示其中发生的各种现象和过程，也都可以借助于动画来实现。

所以动画在教学领域的应用可谓是如鱼得水，可以大有一番作为，它一改以往老师上课时单纯说教、学生课后死记硬背的教学学习模式，而是把枯燥的知识溶于游戏、动画这种轻松形象的方式之中，寓教于乐，使学生在潜移默化之中获取了知识，增长了才干，使学习由负担变成了一种享受。

一、计算机是如何制作动画的

对于动画诸位肯定不会太陌生，也许有的朋友马上想到了许多著名的卡通片和卡通英雄、卡通明星，比如可爱的米老鼠、唐老鸭，勇敢的小老鼠舒克和贝塔，金刚葫芦娃，“齐天大圣”孙悟空以及狮子王辛巴等等，一下子可以列出一长串。这一些都是用传统的卡通片制作法创作的优秀卡通片，不过我们现在要介绍的则是计算机动画，它有自己的特色。当然随着计算机动

画技术的发展，有些著名的卡通片厂商也开始使用计算机制作卡通片，这主要是因为计算机动画能大幅度地降低动画片制作的工作量和成本，而且计算机动画还能产生出某些传统方法所无法做到的效果。其实动画片除了一般常见的卡通片之外，还有一类，那就是木偶动画。但是传统的木偶动画拍摄时是由人控制木偶运动的，不仅动作显得生硬，而且往往造型不太“真实”，一眼就可以看出那只是木偶。由于计算机的介入，已彻底改变了这种状况，现已可以产生出以假乱真的形象来。比如影片“侏㑩纪公园”中的恐龙形象，如果你不知道恐龙早在70万年前灭绝了的话，肯定会信以为真，以为目前世界上还真有这种吓人的怪兽存在呢，或者真像该片开头所说的那样，以为现在可以用生物技术把恐龙“造”出来了呢。而且用计算机生成的这种三维形象不仅看上去很像是“真”的，而且即使活动起来看上去简直也是“活”的无疑。

那么计算机到底是如何制作动画的呢？在回答这个问题之前先让我们看看传统的动画创作过程，因为事实上计算机动画在原理上与其有相同之处。

大家也许已知道电影是由电影拷贝通过放映机放映出来的。而拷贝则是由一格一格分开的静止画面连接起来的，这每一格画面就称为一幅。当放映机以每秒24幅的速度放映拷贝时，拷贝上的静止图像就成了银幕上的活动画面：人就像平时那样说话，并可以随意活动；火车可以呼啸着前进，汽车可以风驰电掣地跑，火箭拖着长长的火焰奔向莽莽的宇宙……

静止画面到活动画面，这其中的奥秘在哪儿呢？原来这是

利用了人类视觉暂留的特性。也许大家有过这样的体验，当我们在黑夜中看到远处有一个在抽烟的人拿烟的手在空中挥了一下时，我们可以发现烧红的烟头在空中划出了一条红色的孤线，而不是看见烟头在动，这就是视觉暂留从中作祟的缘故，因为我们的眼睛有这样一种特性，即在一个景物进入我们的眼睛被我们看到之后，即使它马上消失，它留在我们眼睛中的像却并不会马上随之消失，还能存留一段较短的时间，这就是视觉暂留特性。于是当烟头快速移动时，尽管烟头在每一位置的停留时间极短，但由于视觉暂留，我们就把它与下一位置的像连在了一块儿，于是烟头的轨迹看上去就成了一条连续结的弧线了。电影画面由静到动也正是利用了人类视觉的这种特性，即人眼把相邻几幅画面连在了一起，而这相邻的几幅画面往往就是一个连续动作的“分解动作”，就这样，由于人眼的平滑作用，使银幕上的形象一个个栩栩如生地活动起来了。

动画的基本原理与电影是相似的，也是通过快速移动静止画面产生动感的，只是动画的静止画面不是用摄影机对着演员和真实物景拍摄下来的，而是由人手工或借助于计算机生成的。

例如一组溜冰运动员收回左腿的动画片段。在1号画面上（或称为帧）他开始收回左腿，直至8号画面完全收回，2－7是其间的过渡画面。如果由1号至8号依次播放（比如按每秒24帧的速度），就可以欣赏到运动员溜冰过程中迅速收回左腿的连续过程。

在动画片中，一般把这种一格一格的动画分为两种，一种称为关键帧，它是反映一个连续动作开始和终止的画面，如前例中

1 号帧和 8 号帧，有时为方便也可能把 5 号帧也作为关键(或称次关键帧)，即关键帧定义了一个动作过程的框架。另一种就是中间帧，它是连接两个关键帧的过渡画面。

动画的制作要分好几道工序，工作量是比较大的，最主要的几步就在于画面的创作。当然为创作一部动画片，首先要有一个文学剧本，以及按此剧本设计的分镜头剧本。第二步是角色设计，色括形象设计、服装设计、道具设计等。完成第二步之后，就要开始正式创作画面了。首先创作的是关键帧。顾名思义关键帧是比较关键的画面，中间画面就是按它们来画的，所以关键帧常要求创作水平比较高经验比较丰富的动画师来完成，常体现了动画师的创作水准，是一项创造性的劳动。创作中间画相时来说对技术要求就不那么高了，但由于其画面众多，工作量却非常大，常需要由众多的工作人员在较长的时间才能完成。比如以每秒移动 24 格为例，10 分钟的动画片其需要 24×10×60＝14400 格画面，这其中绝大部分是中间画，工作量之大由此可见一斑。

做完上述工作之后，就要对画面进行测试和修改，这经常要反复几次直至满意为止。然后就要把铅笔稿描在胶片上，这称为线描，描完后还需上色，无疑这两项又是简单的重复劳动，工作量也是比较大的。动画制作的最后几步就是摄制成片，配上声音和文字，有时也许还要剪辑一下，所有这些工作都是每一部动画片所必须要做好的，来不得半点疏忽。每天下午当你坐在电视机前欣赏固定的动画片节目时，不知想到过没有，制作这么短时间的一集动画片竟要经过这么多道工序，花费这么多时间。

计算机动画的出现，改变了动画手工制作这种落后的面貌，使动画师可以从繁重的简单劳动中解放出来，把更多的精力集中在创作上。不仅如此，计算机动画的出现，还为动画师提供了一些全新的创作工具和思路，使他们能创作出许多变幻莫测的画面来。

按其创作原理，计算机动画可以为两种：一种是二维的计算机辅助动画，另一种是基于造型的三维计算机生成动画。另外按动画的记录方式也可分为逐帧动画和实时动画。逐帧动画的原理跟传统的动画相似，计算机逐帧快速放映预先存好的动画画面产生动画效果；而实时动画不需要预先存好每一帧画面，预先存储的只是产生动画画面的程序，播放时由处理器自动执行这段动画程序，产生活动画面（一般是通过各种复杂的变换产生活动效果），所以其动画画面是由处理器实时产生的，故称之为实时动画。实时动画多见于电子游戏，尤其是专用游戏机。实时动画是由处理器实时产生的，所以对处理器的速度要求比较高，尤其是对于产生逼真的动画画面更是如此。目前限于处理器速度，实时动画尚嫌粗糙，不如逐帧动画逼真好看。但随着速度更快的处理器的问世，相信其效果会有很大提高。下面我们主要介绍一下逐帧动画。

逐帧二维计算机动画的制作过程与传统卡通片的制作过程相似，也大致可分为上述几个步骤，只是它是在计算机上借助于动画制作软件完成的，其具体制作过程涉及到具体的动画制作工具，这里不再介绍，下面只想说明二维动画制作中人与计算机的不同分工。

计算机动画的出现，使人们告别了动画制作的手工时代，迎来了制作自动化，大大减轻了人们制作动画的劳动强度。但是计算机也非是万能的，最致命的一点是它不具有像人那样的创造能力（某种程度上这也许是一件幸事，否则人就什么都不如机器了）。所以对于二维动画，颇具创造性的关键帧的创作计算机就无法自动生成，还需由动画师亲自在计算机屏幕上设计好，再由计算机据此来自动生成中间帧。计算机对于重复劳动则是很擅长的，所以中间帧的生成，为画面上色（线描已不需要了，因为画面本身已在计算机中了）等均可由计算机自动完成。人们要做的只是指导性的工作。比如规定运动路线、中间帧的帧数、画面的颜色设置等等。而这都可以以人机对话的方式通过设定各种参数来实现，余下的工作就由计算机自动来处理了，而且处理结果可以马上显示出来供您参考，以判定是否合乎要求，这比起手工制作轻松多了。而且对于诸如上色等工作，计算机可能比人干得更好，因为计算机上色绝不会出现诸如上色不匀、颜色有变化、反光、开裂等人工上色时容易出现的问题。由此可见，在二维动画的创作中，计算机只起了辅助工具的作用，所以常把二维计算机动画称为计算机辅助动画。

如果说二维动画是卡通片的话，正如前面所说的，那么三维动画就有点像木偶动画了。制作木偶剧需要设计木偶形象、道具、布景等，并把其活动过程拍摄下来。同样制作三维动画也需要在计算机上设计好角色、实物和景物，并使各个对象活动起来。也许可以这么说只有计算机三维动画才算是真正的计算机动画，才算真正发挥了计算机动画的优势，当然计算机二维动画

也有其独到的优点，比如处理灵活，可以做得非常夸张、幽默等，但要说真正的计算机动画则非三维动画莫属，因为三维动画从对象的造型设计、空间运动，以至于摄影打光等几乎所有制作过程全由计算机包揽了，这也是为什么把三维动画叫计算机生成动画的原因所在。

在创作三维动画时，计算机首先根据输入的数据，进行复杂的运算，生成模型的三维框架图，然后根据动画师设计好的运动轨迹，自动实现三维实体在三维空间中的自动活动，比如移近移远、转身弯腰、旋转变形、上下跳动等等，运动过程几近乱真。更绝的是，计算机还可以为三维形体制作上一层“皮肤”，显示出各种实体特有的质感。比如电影“侏罗纪公园”中的恐龙形象，据说光一头恐龙就为其“植”了三层“皮”，一层是基本的皮肤，第二层是恐龙皮肤上特有的突起、肿块等，最后一层就是表面的皱纹、污物等，据说使用了多台工作站才完成的。当用计算机为实体披上其特有的纹理外衣，使其看上去更加接近真实形象外，如果再像“侏偠纪公园”那样把它与真实拍摄的场景合在一起，那么不指出来的话，你肯定分辨不出谁是真实拍下来的谁是计算机生成的，这种效果才真真体现了计算机三维动画的“魅力”所在。

二、计算机动画的应用

计算机动画从诞生之日开始，似乎就与影视业结下了不解

之缘，而且目前影视业仍是计算机动画的最大市场。据过去的数字统计，美国在1991年计算机动画产值还只有70亿美元，而到1992年产值已猛升至230亿美元，而且直至今日，计算机动画仍是好莱坞电影制作人津津乐道的话题。从最初的计算机卡通片，到现在运用计算机三维动画制作的科幻片、喜剧片比比皆是。

第一部计算机动画制作的动画片出现在20世纪70年代，片名叫"奥林匹亚"，据说曾长达3分钟。此后随着技术的发展，计算机被频频用于动画片中，大规模的商业应用则始于20世纪80年代。著名的"星球大战"一片中，就有许多画面采用的是计算机动画。另外曾用与制作"星球大战"一片中动画画面的动画制作系统Pixar，曾被用于制作另一部计算机动画片"锡玩具"(Tin Toy)，该片曾获得了奥斯卡动画片奖。目前计算机动画更多地被用于科幻片中，创造了一系列令人目瞪口呆的科幻形象。看过"终结者Ⅱ"的人肯定还记忆犹新，来自外星球的液态金属机器人T－1000竟然能通过变形渗入直升机内部再恢复成人形；面对熊熊烈焰竟毫不犹豫地直穿过去；当在精神病医院警察来抓他时，他又能通过迅速变形化成一张地毯，弄得警察大惑不解；最后他"壮烈"地被一团火红的钢水吞没。这一系列超乎想象、神奇无比的精彩画面都是计算机动画的杰作，如果用传统的模型来拍摄，效果绝不可同日而语。而不久前又刚创下票房纪录的"侏俘纪公园"，里面除了运用造型之外，也使用了不少计算机动画，据称其中最激动人心的6分钟用的就是计算机动画。大家也许还记得在广阔的草地上，恐龙由远及近疯狂地追

赶博士一行人的精彩场面，画面上由计算机生成的恐龙简直就是跟“活”的一样，跑得惟妙惟肖，扣人心弦。

计算机动画的另一巨大应用市场要算电视片头和广告制作了。我国在亚运会期间推出的“亚运之光”电视节目，就首次采用了计算机动画技术。那五彩缤纷、绚丽多彩的画面给人以耳目一新的感觉。中央台新闻联播节目的片头也从 1991 年元旦开始采用计算机动画生成。另外用计算机动画创作新颖别致的广告也大有前途，计算机动画高效的创作过程和别具一格的创作手段，不仅能提高广告设计师的工作效率，还可以藉此设计出能收到意想不到效果的广告作品。怪不得如今的计算机展示会上，参展商也都以特别设计的多媒体演示系统为自己的产品作广告。而这种演示系统目前用得最多的就是计算机动画这种人们喜闻乐见的形式。

正如本节一开头就提到的那样，计算机动画另一个有前途的市场就是教学和培训。当然目前这两个领域计算机动画应用尚不及前两者多，但正如多媒体技术正在开始走向家庭走向学校一样，计算机动画也正在走向这些应用场合。也许有一天当诸位坐在教室里或参加什么培训时，面对的将不再是黑板和粉笔，而是大屏幕显示器，手里拿的也不再是笔和纸，而是操作杆和键盘。

第四节　多彩多姿的计算机视频

这天晚上做完作业之后，杨波又打开电视机收看交互电视网上的电视节目，自然他又选了一部描写去中生代探险的科幻片。看到主人翁被霸王龙追赶只好仓皇跳入湖中时，突然发现画面上闪现出一头他以前没见过的恐龙，鸭子似的嘴巴，头上还挺着个棒球状肉管，也不知作啥用，一会儿又消失在湖中不见了。杨波赶紧拿起遥控器定住了画面，倒回去找到那个画面，把它"剪"了下来，他一时也说不出是什么龙，只好附上自己的问题，连同捕获的画面通过多媒体网络发给了在古生物所工作的叔叔。一会儿叔叔的答复传过来了，他告诉杨波这头恐龙叫棘鼻青岛龙，大概生活在白垩纪时代，就活跃在我国山东一带，它的"鸭嘴"别看它很难看，其实用处大着呢，能把普通食物瞬间磨得粉碎；而头顶上的肉管更是大有用处，关系着它的身家性命，因为遇到敌害时，他就可以跳入水中沉下身子只留肉管在水面上呼吸，这样它就可以生命无忧了。原来那就是它的鼻子，杨波暗暗好笑，心想自然选择的结果竟使大自然创造这么一个有趣的怪物，真有意思，而更令杨波得意的是，它就曾产于我国。

疑问消除了，杨波又打开画面，继续关注片中主人公的命运去了，如果他脱不了身，杨波还打算帮他一把，引来几头角龙跟

霸王龙打上一架。看到这里，也许聪明的读者已有很多疑问了：杨波看的这是什么电视呀？倒像是一种电子游戏。确实不是普通电视，因为他不仅能暂停画面，随便裁剪画面，而且还能与电视节目相互作用，改变主人翁的命运，现在的电视可不行呀，对了，这就是未来的交互式数字电视，而现在我们的电视节目还大都是模拟的，不具备交互功能。数字电视节目现在已经有了，但还仅停留在可以选择节目上，交互作用还不很强，上面杨波看的那种电视还属于明天。但是随着计算机数字视频和高清晰度电视（HDTV）的出现，这一天已不遥远了。本章就谈谈现在的数字视频技术和在多媒体系统中的应用。

一、计算机数字视频的来源

对于电视和录像乃至影碟，相信大家都不陌生了，茶余饭后，看看新闻，看看录像已成了休息的一种方式；而除夕之夜，家人欢聚一堂观看中央台的春节联欢晚会俨然也已成了一种习俗，电视已走进了千家万户。另外稍具电学知识的人也都知道，电视机播放的节目来自电视台发射的电视信号，要不就是录像机或影碟机输出的电视信号，其中与电视画面有关的电视信号一般称为视频信号（电视信号中同时还有音频信号，同步信号等）。

计算机处理的视频信号原始的来源也是上面几种，即摄像机、录像机、电视机和影碟机等。不同之处在于计算机处理的是

数字视频，而前面几种设备处理的视频信号都是模拟视频。那么模拟和数字到底是什么意思呢？下面谈谈模拟信号与数字信号的区别。所谓模拟信号指在时间上和幅度上都是连续的信号，因此对于模拟信号来说不仅信号的一个幅度值可能有无穷位长，而且在一段时间内其幅度值也有无穷多个，因为时间 t 也是个连续量。很明显这样的信号计算机是无法处理的，因为计算机存储器的字长和容量总是有限的，况且也不可能无限止地运算下去。所以为了便于计算机处理，模拟信号必须数字化，对于视频来说，就是要把模拟视频转化成数字视频。转化的过程可分为两步，第一步是在时间上把信号离散化，即把一段视频信号仅每隔一定时间才取一个值（称为取样值）；第二步是把这有限个取样值再在幅度上离散化，表示成有限位 0 和 1 的排列组合，这就变换成了数字信号。这两步分别称为采样和量化编码，合在一起称为模/数转换。也许大家有疑虑，这样处理之后，电视图像不会有损伤吗？回答是否定的，理想条件下只要采样速率足够高，即每秒钟采样的点数足够多，那么采样必定可以保证这种处理是无损的，即当把这些数字量进行数/模转换变回模拟量后仍可恢复原来的图像（当然实际情况下损失总是难免的，不过一般人眼是看不出来的）。

现在大家已经知道了计算机处理的数字视频信号（称数据也许更合适些）是如何得到了，那么是不是就可以直接储存这些数据了呢？大家已经看到了静止图像的数据非常大，以至于在存储之前不得不进行压缩，由此可以推测，数字视频也是非压缩不可的。事实正是如此，以把连续画面离散化成 640×480 分辨

率的图像为例，如果每像素用 24 比特位，每秒刷新 30 帧，那么每秒视频图像的数据量为 640×480×24/3×30=27MB，这样大的数据量不用说存储，就连送到计算机都困难，因为计算机内普通总线（传输数据的线路）的传输速率也不到 27MB/S。退一步讲，假如能送到计算机，那么即使存在 1GB 的光盘上，最多也只能存储 1 分钟的视频图像，这样高的成本那又有什么意思呢？

于是首先想到的一个办法就是数据压缩，尽量减少数据量，避免重复存储相同的数据。对于静止画面的压缩在图像一节中我们已谈过了，这种方法自然也适用于活动视频图像一帧之内的数据压缩，另外对于每秒 30 帧的画面，还可以进帧间压缩。大家也许不难想到，对于实际不是高速运动的画面，相邻两帧甚至多帧图像之间具有很大的相似性，这跟动画片中相邻两格画面只有稍微有些变化是一个道理，如果能充分利用这种帧间相似性进行帧间压缩的话，就可以大大减少视频数据的数据量。计算机视频数据压缩正是同时采用了帧间数据压缩技术和帧内数据压缩技术，才使视频数据量达到了计算机可以“忍受”的程度，使其可以通过总线输入计算机内部然后转存于外部存储器上，如光盘、硬盘等。当然恢复视频图像时，也需要计算机进行解压缩运算，这是一个与压缩相逆的过程。下面看看压缩的效果。以著名的 MPEG（Moving Pictures Expert Group）全活动视频压缩标准为例，通过这两步压缩，其压缩比可以达到 100：1，按此压缩比压缩一部一般长度的电视节目或影片就可以把它存储在一张光盘上了。从 1994 年初开始在多媒体产业界备受关注的 VCD（Video CD）视频播放系统，就采用了 MPEG 压缩

标准，一张普通光盘可以存放72分钟的视频节目(包括声音)，这已达到了实用的程度，从去年开始V—CD也陆续走上了我国各家电商场的柜台。当然除MPEG外，还有其他的压缩典范，如DVI、CD—I等，这里就不再介绍了，而把它们放在第六章多媒体系统中再作详细陈述。

二、交互式计算机视频

尽管由于计算机视频数据量庞大，再加上目前处理器速度上的限制，使计算机视频处理技术的发展相对来说慢了点，但从数字视频所具有的清晰度高、抗干扰性强、失真小、处理灵活等优点来看，它的发展又是必然的。尤其是数字视频便于处理的优点，更使其在今后的交互式视频系统中大有用武之地。下面就谈谈一般的多媒体交互式视频系统的特点。

交互式视频一般也融合了其他媒体技术，不仅具有高清晰度画面，还有高保真的数字音响、精美的图文、动画等，当然其中最重要的一点，就是它的可交互性。观看交互式节目不同于观看现在的电影、电视，观众不再是仅限于安静地坐在屏幕前面观看，而是可以任意挑选自己喜欢的节目，并且还可以自由地与节目交互，参与编排节目内容和进行过程，甚至作为一个角色参与其中的工作或演出，使观众更多地参与到节目中去，亲身感受其中的酸甜苦辣，使观众集“导演”、“演员”和观众于一身。不过目前这样的交互式电视网尚没有听说，但利用光盘作为节目载体

的交互式光盘多媒体系统如CD—I已经问世，当然在交互性能上尚有一定距离，相信不久之后，随着技术的成熟，会有大的突破。

三、目前计算机视频的应用

目前计算机数字视频主要被广泛地应用于娱乐业中，各种各样的数字视频播放系统已相继问世，如从去年开始走红我国的VCD系统，不久之后即将推出的DVD系统等，这些系统的共同特点是以光盘为存储介质，由出版者灌好数字视频节目出版发行，用户购买专门的播放系统或在MPC上配置专门的视频解压卡就可以播放这些光盘视频节目了。

计算机视频的另一个巨大应用是与通讯系统相结合，这就产生了人们早就开始幻想的但又迟迟看见其踪迹的视频会议系统、可视电话系统以及银行视频电话储蓄系统等。有了视频会议系统之后，我们昔日坐在办公室里甚至家中开会的梦想就实现了，再也不用为参加一个在异地召开的会议而东奔西跑了；而可视电话的出现，更是使人们多年来通过电话“面对面”交谈的梦想成真。下面就谈谈这方面的情况。

对通话通讯和非通话通讯的研究表明；通讯有效性的55%依赖于面对面的视觉效果，38%依赖于说话者的语音，只有7%依赖于讲话的内容。而且大家也许也早就听说过人是通过视觉获取大部分信息的，俗话说得好“百闻不如一见”嘛！所以现在

普遍使用的电话通讯方式只能部分解决通讯有效性问题。

而将数字视频、数字语音技术与通讯技术相结合推出的视频会议系统则可较圆满地解决通讯有效性的问题。目前，美国利用综合业务数字网、数字交换网等连接视频会议系统，已开展出租业务，只需很少费用即可在相距很远的两地进行“面对面”的谈话。视频会议系统不仅可用于会议、会话，而且还可应用到许多要求“面对面”的场合，如视频远程教学，远程医疗服务，远程招聘人员等等。有资料统计表明，美国一年总的出差费用超过 1500 亿元，一个招聘经理人员的公司 1993 年一年对 9 万人进行第一轮面谈就花去了约 1.35 亿元的出差费，而且占用了大量的人力物力。如果采用视频会议系统则不仅将提高办事效率，还减少了出差时间和费用。

另外，视频通信与控制相结合，还可进行远距离现场监测和指挥，可用于现代军事通信、交通管理、电力管理等方面，使现场信息及时、准确地传给指挥调度中心，使中心可根据现场情况迅速做出判断，并可对现场实施实时控制和指挥。所以，国外有的资料甚至认为视频通信技术将在很多领域引起一场“革命”。

由此可见，视频与通讯的结合不仅造就了一种全新的通讯方式，而且在许多应用场合都有其发挥的余地，那么一般的“视频会议系统”是如何工作的呢？以一般典型的会议形式为例，“视频会议系统”工作时的情形如下：各个会议点的多媒体计算机，将反映各个会场的主要场景，人物及有关资料的图像、图片及发言者的声音同时进行数字化压缩，通过通信系统，发向指定的各个方向；同时，在各会议点的多媒体计算机上，通过数字通

信系统实时接收、解压多媒体会议信息，并在显示屏幕上实时显示出指定会议参加方的会议室场景、人物图像、图片和语音等。

通过上面的简要介绍，大家可以看出，计算机视频由于其特有的优点，不仅可广泛应用于娱乐、教学场合，更可以与通信技术控制技术相结合，广泛应用于生产生活领域，创造出非凡的经济效益，并极大地方便人们的生产生活，必将在许多领域引发一场“革命”。

第四章 能说会唱、善解人意的计算机声音媒体

不知从什么时候起，机器人开始频频出现在科幻作品中。这些智力过人的机器不仅能说会道，还常常勇猛无比，成为作品中主人公的好助手。看过“帝国反击战”一片的人也许还记得其中的几个机器人，尤其是“公主”那个人形机器人不仅脑瓜转得快，嘴巴也厉害，常喜欢发表些意见，以至于后来被敌人发现了拆成了零件。能造出机器人来延伸人的能力一直是人们的梦想，理想中的机器人应在某些方面智力过人，并能用语言与人直接交流，可惜的是限于目前人们对人脑的了解，像科幻作品中智勇双全的机器人目前还造不出来，不过与机器用语言直接交流的梦想正随着语音处理技术的飞速发展而成为现实。语音处理技术尤其是语音识别技术近年来发展很快，多种试验性的语音识别系统已经问世，有的还成了商品，同时语音合成技术也取得了很大成果，这就使得直接的人机对话有望在不久的将来成为

现实。

语音信号处理技术是声音媒体技术的一个重要领域，声音媒体的其他方面还有音乐、声响等音频信号处理。对一般音频信号的处理常涉及压缩编码、储存、恢复、合成等。比如通过音频合成计算机可以产生出许多你闻所未闻的声响效果，比如机器人金属般轻脆的说话声、外星人好像从肚子里发出来的声音等各种稀奇古怪的声音。音频合成的另一个重要应用就是用于模仿各种音色的乐器去演奏音乐，并可把用户提交的乐谱自动转化为音频信号。另外计算机还可用于播放各种高保真数字音响节目，甚至还可让你不仅仅满足于被动地听，而去让你自己谱曲由计算机为你演奏，并且还可让你为音乐配上活动画面，犹如MTV，使你在欣赏各种风格的音乐作品之余，还可领略一下自创歌曲的乐趣，使你拥有一个可以尽情发挥音乐天赋的地方。

计算机声音媒体除可应用于上面的人机对话，声音娱乐等领域外，还可广泛应用于日常学习生活及生产领域。初学英语的人都知道若无老师在旁指导自己学起来是较费劲的。不说别的，单单读音一关就比较难过。每当碰到一个不会读音的单词时，如果会音标还可以查查字典（且不论能否拼得准，说得地道），如果不会音标又没人帮你，那就只好对着单词干瞪眼了。现在有了英语语音词典就好办多了，因为它不仅能当普通词典使，而且更重要的是，它还能为你朗读单词的读音，而且读得字正腔圆、十分地道。拥有这种语音词典你就相当于随时有个外籍教师陪伴着您了。现在各种会“说话”的语音词典、快译通、电子记事簿已经面市，走上了许多商场的柜台。计算机语音控制

还可以应用于危险场合的控制，人只要在安全的地方通过话筒“发号施令”就行了，里面的机器就会按照您的吩咐自行操作。至于里面的状况则可通过摄像机、专用传感器输送出来供你判断。同样，在家中，也可以运用计算机语音技术声控各种家用电器如电视机、洗衣机、微波炉等使它们一个个都变成了“听话”的好帮手。

看到计算机声音媒体能为我们干这么多有趣又有用的工作，心急的朋友可能已在心中嘀咕了：为什么声音经计算机处理后就变得这么有用了呢？它们到底是如何工作的呢？不忙，不忙，切勿着急，计算机声音处理是一个涉及多门学科的交叉学科，涉及的技术很多，且等我慢慢介绍。

第一节　数字音频的一般处理

对于声音，大家是再熟悉不过了，我们这个世界几乎没有绝对无声的地方，小到蚊子的嗡嗡声，大到雷电的隆隆声；有清脆响亮的敲锣声，也有深厚低沉的擂鼓声，小提琴奏出悠扬的乐曲，钢琴弹出铿锵的音符……而我们人自己发出的声音更是感情丰富，多姿多彩，比如高兴时的欢笑声，悲痛时的啼哭声，生气时的嘟囔声，失望时的埋怨声……

那么声音是如何产生的呢？学过物理的朋友很轻松地就能

回答这个问题:声音是由空气的振动产生的。例如当天空中闪电相撞时,巨大的电能倾刻间转化为了热能而热量又加热了周围的空气,使空气迅速膨胀,并向四周扩散引起空气的复杂运动而产生了巨大的雷声。另外大家也许还记得,声音是以声波的方式传播的,这在我们的弹吉它时从弦的振动形状上可以观察到这个现象。声音经话筒拾取之后就被转换成了电信号,这时就称为音频信号。

音频信号表现为幅度随时间变化的连续波形,为此我们引入描述信号特征的两个重要的基本参数——幅度和频率。幅度表示声音音量的大小,而频率则表示信号变化的快慢。如果声音的幅度太小,则因其强度不够,限于我们人耳的灵敏度就可能听不见这个声音,所以只有达到一定强度的声音我们人耳才能听得见。另外若声音变化太慢,频率太低,比如按小时或天数变化的大气压,我们也不易听到;但如果 声音变化太快,即频率太高,由于我们耳朵反应不过来也听不到。所以人耳能听见的声音实际上只是自然界很少的一部分,一般人能听见的声音频率在 20Hz～20KHz 之间(1Hz 指每秒变化一次),而人平时说话时的声音频率的宽度则比它还窄,仅在 300Hz～3000Hz 之间。另外女性的声音频率普遍地比男性的高,这就是我们为什么说女孩是“尖嗓门”,因为声音频率高了,我们听起来就比较“尖”。

对一般声音信号的分析表明,声音信号实际是由许多不同频率的信号组成的,这称为复合信号。单纯只有一个频率的声音信号就是正弦波,所以一个复合的声音信号在数学上可以表示为许多幅度和相应不同的正弦(或余弦)波的和。下面再介绍

几个有用的与频率有关的概念。我们在听收音机时，常会听到波段的概念，一个波段实际上就指从一个频率下限到一个频率上限的频率段，频率段的长度就是频带宽度，简称带宽。我们上面说了一般人耳能听见的声音频率在20Hz～20KHz之间，那么这段声音的带宽就是20KHz－20Hz＝19980Hz≈20KHz。事实上，一个我们能听到的声音除了在20Hz～20KHz有能量外，还往往具有高次谐波，即在20KHz以上还有能量，只是我们人身听不见它们而已，但如果用话筒记下来的话就可以发现这些高频信号，但由于对我们人耳没什么作用，在声音处理时就往往要把高于20KHz的音频信号滤掉，这就要用到滤波器。滤波器就是把信号的一部分频率滤掉而保留另一部频率的设备，这样看来，我们人耳天然就是一个滤波器了。

正如我们在视频中提到的那样，经摄像机记录下来的视频信号是模拟信号，即在时间上和幅度上都是连续的，而计算机只能处理二进制的数字信号，所以需要把模拟视频轻换为数字视频，这称为模/数转换（A/D转换）。同样由话筒记录下来的音频信号也是模拟信号，为了便于计算机处理，也有个模/数转换问题，把模拟音频转换为数字音频。

模拟音频的数字化原理同视频的数字化原理是一样的，分为时间上离散化和幅度离散化两步，具体过程请参阅视频部分，这里不再赘述，下面仅指出音频数字化与视频数字比之间不同之处。

前面一章已经说了，为了使数字化后的信号在恢复时不失真，要求每秒采样的点数要足够多，即采样速度要足够快，事实

上，采样定理指出，这个速度的下限就是信号上限频率的2倍。所以20Hz～20KHz的音频信号数字化时采样速率至少在$20\times10^3\times2=4\times10^4$点/秒以上。我们前面已经提到，20Hz～20KHz是人耳的听觉范围，但并没说人们听到的声音带宽一定就有这么宽。事实上，我们平时听到说话声就只有300Hz～3000Hz带宽。所以现在提出这样一个问题：对于具体一个声音，到底应取多大的带宽？这个问题是很重要的，如取窄了，声音要失真，取宽了则采样速率就高，数据量就大了。实际上，这个问题的回答涉及到了数字音响系统中用带宽来衡量声音质量的问题。

研究表明，音频信号中低于200Hz的低频信号只是用来增加话音的自然度和说话者的风度，而高于7000Hz的高频信号只是同来提高话音的可懂度和增加对多种摩擦音的区分能力，所以对于一般的语音理解，200Hz～3400Hz的带宽完全足够了，这正是目前长途数字电话上使用的语音带宽。而对于高保真音响，显然这样的带宽是不够的，因为音乐作品中包含有乐器奏出的高音，另外，男、女高音歌唱家唱歌时的带宽肯定比我们平时说话声的带宽要高，于是对于超级高保真(Super Hifi)音响系统，带宽就要取为人耳能听到的整个声音频段，即20Hz～20KHz，另外考滤到实际实现时滤波器的过渡特性，采样率就一般取为44.1KHz，比20KHz的2倍略大一些。这是最高的声音质量等级了，常称为CD等级，而数字电话用的是最低的等级，其间还有两级声音等级，即50Hz～7KHz的AM调幅无线电广播质量和20Hz～15KHz的FM调频无线电广播质量，分

别应用于相应的广播系统中。

确定音频带宽之后，就可以确定采样率来对模拟音频信号进行采样。比如确定某信号只要话音质量就够了，于是用滤波器滤出200Hz～3400Hz的音频信号，再用不小于二倍最高频率的采样率比如8KHz对信号进行采样。采样后的数据还要进行量化编码处理，即把音频信号在幅度上进行离散化，并用有限位0和1对其进行编码，转化成计算机能处理的数字信号。对于一般音频信号而言一般可用16bit或8bit进行量化，区别在于同8bit可将幅度分成256个相等的单位，而用16bit可将其分成65536个相等单位，显然后者能把一个信号在幅度上分得更加细致，从而可以更接近于原来的音频波形，保真度也就越高，但同时数据量也增加了一倍。

提到数据量，我们不得不指出，跟视频数据相似，音频的数据量也很大。比如数据量最大的数字立体声超级高保真音频信号，每个通道每秒就有44100个样本，每个样本值再用16bit即两个字节来量化，则每秒钟有44100×2×2＝176400字节，存储1小时这样的音频得需要176400×3600＝635MB，即大概一张光盘的容量。这样大的数据量对于高保真音响来说也许是值得的，因为它就是获得这种高保真效果，但对于话音通讯就没必要了，因为话音通讯跟它不一样，话音通讯的目的是传递信息，只要对方能听懂我的意思就行了，最多也只是让对方能听出是谁在讲话，即保持一定的自然度就可以了。所以对于话音除带宽取的较窄，量化比特数用得较少外，还可采用各种压缩算法来设法压缩数据量。事实上不仅对于话音通讯、语音的在压缩编码，

对于以后的各种处理也是有益的，比如文本——语音转换和语音识别中要存储大量的标准语音数据，对语音数据进行压缩自然可以节省许多存储器，而且也可使查找速度加快。

对于音乐，除了对其通过话筒记录下来的音频信号直接进行采样、量化、编码外（这一般称为波形编码），还有另一种面向音符的编码存储方式，也用得十分普遍，而且这种方式也能大大压缩数据量，减少存储空间。不过，在详细介绍语音和音乐的各种编码方法之前，我打算先介绍一下一般音频信号的处理。

计算机对音频信号的处理除了音频的捕获和数字化外，还有数据存储、音频编辑和音频播放等处理。由于音频数据量大，实时性强，这就要求计算机的处理速度要很快，目前一般的计算机对此要求是有困难的，普遍的办法是外插一块带有高速数字信号处理能力的 DSP（Digital Signal Processor，即数字信号处理器）的声效卡（详见多媒设备一章），由声效卡来完成音频处理时实时性要求较高的部分，如模/数转换、压缩存储、音频重放等。

与视频相似，音频数据的存放一般也用光盘，而且采用特定的文件结构，以区别于一般的程序数据文件和图像数据文件。由于计算机处理数字音频的灵活性，使人们在编辑音频时可以进行各种方便的处理，如去除开头和结尾的噪音段、无音段或中止段，而且还可以重新安排或删除音频数据，甚至把一段话音加于背景音乐上，把几段音乐连成一长段再加上各种音响效果等等，可让你自由裁剪，随你优化组合，有时甚至能造出出乎意料的声音效果来。

第二节　MIDI合成音乐

玩过电子琴的朋友也许还记得，电子琴与其说是“琴”还不如说成是键盘乐器，因为它除了能模仿钢琴、手风琴等“琴”的音色演奏旋律之外，还能模仿笛、管等乐器，而且更与一般琴不同的是，它还能为你加入鼓点，加进去各种节奏，简直就像是一个小乐队。那么电子琴这么多奇妙的功能是如何实现的呢？

我们前面已经说了，一个声音是由许许多多的频率分量构成的，所以对于特定的声音如果我们能分析出它的多个频率分量，然后据此用电路生成多个频率的分量并把它们合成到一块儿，于是我们就能模拟出这种声音了。另外我们也知道藉以区分各种乐器的音色在物理上表现为一个基音和多个泛音（高次谐波）的组合，于是如果我们能用电子合成器来合成一种乐器的基音和泛音的话，那么我们就能模仿该种乐器的音色了。

但仅是这种简单的合成效果并不好，为改善合成音乐的真实感，音乐家意识到必须把许多合成器连接起来，以产生出复调和多音色声音。复调声音是指同时演奏若干音符时发出的声音，如钢琴、吉它等乐器演奏和弦时就要同时奏响几个音符，而双簧管则不能；多音色是指同时演奏几种不同乐器时发出的声音，这有点像乐队演奏。为了实现合成器连接，就需要规定各种

合成器之间的接口标准，这就是目前常说的 MIDI(Music Instrument Digital Interface，简称 MIDI，即乐器数字接口)标准，它不仅规定了乐器之间数据传送的约定，而且还定义了与计算机的接口，需要说明的是这里所说的乐器一般都指电子合成乐器。

我们不想在这里过多地涉及乐器之间及乐器与计算机之间的硬件连接问题，只想谈谈计算机是如何通过 MIDI 接口控制合成器演奏音乐及特殊声响的。首先乐器通过与计算机的接口把同音乐有关的信息主要是乐谱及乐器信息传给计算机，计算机收到之后经过适当处理，以 MIDI 文件的格式存放在存储器中(当然也可以直接把信息输出给另外一台合成器去演奏)。一个 MIDI 文件是存储记录 MIDI 信息的标准文件格式，它包括有音符，定时和通道选择等指示信息，而且对于音乐的每一个音符还具体定义了演奏该音符时要用到的有关信息，如关键字、通道号，该演奏的时间长度，奏响时的音量和演奏速度等。如果计算机又把 MIDI 文件存储的音乐信息输出给合成器，合成器就能根据 MIDI 文件奏出相应的音乐来。

也许有读者要问：前面不是已经说了音乐可以通过波形数字化来存储吗，干吗还要使用 MIDI 呢？这主要同为 MIDI 有下面几点好处。首先因为 MIDI 是一种针对特定音乐，而向音符的控制语言，MIDI 文件存储的主要是音符及定时信息(这很象乐谱)，所以 MIDI 文件占用存储空间较少。比如一段持续 1.8 秒的音频波形若以 22.05KHZ 采样、8bit 量化的话要占用 42KB 空间，而一个典型的持续 2 分钟的 MIDI 文件只需 8KB

存储空间，压缩率是很可观的。MIDI 的第二点好处是编辑方便。这是因为对于音频波形编辑，我们看到的仅是其波形，虽说直观，但因我们并无解读波形的能力使我们想象不出这是什么声音，除非预先听过。而 MIDI 文件因为是面向音符的，修改起来犹如修改乐谱一样，这对于懂乐理的人来说是非常方便的，而事实上有的 MIDI 文件编辑屏幕上，用户看到的就是一本五线谱，所以在这样的环境下编辑音乐实际上犹如作曲一般。与一般手工作曲不同的是，作完一曲甚至一段之后，就可以让计算机马上为你演奏，看看效果如何，你再不用为不会弹钢琴或弹不好而发愁。看来，MIDI 音乐真是作曲家梦寐以求的好帮手，即使我们这些不懂作曲而又对作曲颇有几分好奇心的人，也可以一试身手，听听我们自己创作的稀奇古怪的“宇宙音响”。MIDI 的最后一个好处是同一段音乐通过计算机的指派可以让不同的乐器来演奏，甚至让多种乐器同时演奏，并且还可以加入节奏、鼓点，俨然就是一个小乐队在演出。所以 MIDI 通常多应用于计算机音乐界，用于演播室控制和音频产生，所以很有可能有一天你听到的一段奇妙的音乐就是 MIDI 音乐呢！

第三节　语音的压缩编码

我们在前面已经指出，音频信号数字化后的数据量也是很

大的，即使是要求较低的话音音频，若以 8KHz 采样，每样点用 8bit 量化的话，也需要每秒 64K bit 的数据量，这对传输信道容量和存储容量是一个挑战，要知道即使现在最快的调制解调器其传输速率也不过 20 多个 Kbps。为了解决这个矛盾，跟视频数据一样，人们也想出了各种各样的语音压缩技术。

语音的压缩编码方法一般可分为两种，即波形编码法和分析——合成编码法，两者的原理和所依据的标准是不一样的。波形编码法以尽可能地保持语音波形形状为标准，同时采用各种方法使编码后的数据量尽可能地少。这是一种信号进行压缩编码的通用方法，而不仅仅限于语音信号，一般的音频信号和视频信号都可以用。而分析——合成法则主要是针对语音信号进行压缩编码的，因为它所依据的语音生成模型就是根据人的发声机制构造的。它根据输入的语音数据提取出语音产生模型的参数，再对这些个参数进行压缩编码；而译码时就用这些参数去驱动语音产生模型生成合成语音。合成语音以尽可能地保留原有语音的语义为前提，相对波形编码恢复的语音，其质量是比较差的，但因它只需对参数进行编码，所以可以大大压缩数据量，对于这种编译码系统一般称为声码器。

我们在前面提到过，一个信号既可以在时间轴上用波形来表示，也可以在频率轴（或称频域）上用各个频率分量来表示，事实上这种时域波形与频域分量表示是一一对应的，所以对语音信号的波形编码既可以直接在时域上进行，也可以先把时域波形变换成频域表示，在频域进行编码。当然对于后者，解码后首先得到的是波形的频域表示，需再作一次频域到时域的变换，才

能恢复成原先的时域波形，这种波形编码方法也称频域编码，如常见的子带编码、变换域编码等，其好处是压缩率比时域编码高。

下面先介绍波形的时域压缩编码。大家知道，我们说一句话时常常是有上下文联系的，所以有时即使我们没听清一句话中的某几个音，也照样能理解其中的意思。波形的时域压缩编码就是利用了语音的这种前后联系。语音中联系最强的大概要算相邻的两个采样数据了，这是因为信号的能量总是有限的，瞬间不可能发生很大的突变，比如前一样点数据值很小以至于接近零而其后一个样点突然增至最大幅度值，这种情况出现的机会极小以至于我们可以认为它是不存在的，所以说语音信号一般情况下变化是缓慢的。利用语音这种特性，我们就可以不直接存取样点数据本身，而只存取它与前一样点值的差，这样数值就小得多了(因为它们是缓变的，幅度相差不大)，于是可用很少几个比特位来对它进行量化编码了，比如 4bit，这样不仅数据压缩了一半，波形的失真可能反而也小，这种压缩编码方法常称为差分脉冲编码调制(DPCM)。如果说大家对于仅仅利用前面一个样点的值来做减法还不放心，那么我们干吗不多利用前面的几个样点值呢？回答是肯定的，这就是自适应差分脉冲编码调制(ADPCM)的基本思想，ADPCM 利用了前面 M 个样点的值，根据对现在这点影响的大小乘以加权系数之后相加(这称为预测)，再用这个预测值去同现在这个样点值作差，最后对这个差值来编码，这就是它的基本过程。由于 ADPCM 利用了前面多个样点的信息，所以波形失真自然比 DPCM 小。现在已有专门

的ADPCM标准，这就是国际电报电话咨询委员会推荐的G.721ADPCM标准。

波形的频域编码方法也很多，这里主要介绍子带编码和变换域编码两种。从前面几种压缩方法看，其基本原理都是从减少输入语音信号的冗余度（相关性）着手的，只是它们使用的是时域预测的方法，而子带编码用的则是分隔频带的方法来去除冗余度。它首先将输入信号分割成几个不同的频带分量（称为子带）然后，根据不同子带的实际情况合理地分配量化比特数，并使由量化产生的噪声适应人耳的听觉特性，获得更好的主观听音质量。比如对语音质量影响较大的基音和共振峰都集中在低频段，它们要求比较高的精度，我们可以为它们分配较多的量化比特数，而高频段则可以相应少分配几位。于是经这样的优化处理，不仅降低了数据量，而且获得了较好的主观听音质量，以致于不仅语音用它来压缩数据，甚至连保真度要求非常高的CD等级的音频也可用它来压缩数据，比如国际标准化组织（ISO）关于CD等级的20KHz带宽128kb/s声频编码标准，就采用了子带编码方法。与子带编码类似，变换域编码也是用划分频带的方法来去除冗余度的，只是子带编码的划分方法是把整个语音频带先划分成2～4个具有一定带宽的子带（即使宽带音频也不过32个子带），然后才对每一子带再进行采样（采样速率跟子带带宽有关）和量化编码；而变换域编码则干脆多了，它直接用离散余弦变换把整个频带进行离散化，一般用N个频率分量（N一般为128～256）来代替，然后再对N个频率分量进行量化编码。所以说变换域编码是直接把一个连续的频带用N

根谱线或称频谱分量来代表了。变换域压编码的效果是惊人的，如果采用合适的变换类型，合适的 N 值，量化比特数分配也恰当的话，变换域编码能在平均每样点 2bit 的速率（原始语音是每样点 8bit）上给出高质量的重建语音，甚至在每样点 1bit 时仍能得到相当好的语音编码质量。

大家也许已经发现，波形编码为了最大限度的重现原始语音，它的压缩比是不很大的，即使效率较高的变换域编码最大压缩比也不过 8：1，但它的好处是失真比较小，语音的自然度也比较高，多见用于数字电话。而我们下面要讲到的各种声码器编码方法与此相反，它可以大幅度地降低数据量，达到很高的压缩比，但同时也使合成后的语音失去了诸如自然度等很多特性，以至有的情况下即使是熟人我们也无法从其合成语音中辨别出来，当然，听懂是没问题的，否则该方法就没意义了。举个例子，美国国家安全局（NSA）使用的保密通信系统，采用 1992 年公布的线性预频编码方案的话，数据速率每秒只有 2.4K bit（可记为 2.4kb/s），比起压缩前的速率 64 kb/s，整整压缩了 30 多倍，但其话音质量失真很大，平均意见得分（亦称 MOS 分，是一种主观评测语音质量的方法）仅得 2.5 分（满分是 5 分），低于一般的通信质量（MOS 分 3.5 分左右）。而用 NSA1989 年公布的激励线性预测声码器编码方案的话（简称 CELP），则 MOS 分可达到 3.2 分，已接近通信质量，数据速率也可以，可达 4.8kb/s。那么这些编码方案何以能达到如此高的压缩比呢？对这个问题的回答涉及到了人是如何发声的问题。

通过多年的研究，现在一般都认为语音是由肺部产生的气

流经过声道时经声道调制生成的，即由于声道的各种形变，改变了气流各部分的压强和速度，最后经嘴唇辐射出去时就形成了各种不同的声音。人的话音尽管听起来五花八门，有多种不同的音素，但基本可以分为浊音和清音两类，这是根据激励声道的气流的波形不同来区分的。

浊音表现出一定的周期性，而清音则杂乱无章毫无规律可言，犹如随机噪声一般。确实，从两者的激励脉冲也可以看出这一点，浊音是由声带推动产生的准周期脉冲激励声道产生的，而每一次振动都使一股气流从肺部流进声道，各股气流之间的间隔时间就称为基音周期；清音则不同，它是随机的空气流通过声道时，受声道某些部分压缩引起的，表现出一定的随机噪声特性。另外对于跟浊音有关的基音周期还想指出一点，即基音周期“男女有别”，男性的基音周期一般从 5ms 到 20ms，而女性的基音周期相对要短一些，从 2.5ms 到 10ms，即女性的基音频率较高。

我们前面已经指出，语音的合成——分析编码是基于语音的生成模型的，现在我们已经介绍了一般的语音发生机制，下面就介绍常用的语音生成模型。

这个模型很好地体现了上述语音生成原理。周期脉冲发生器作为产生浊音的激励源，其基音频（基音周期的倒数）由外界输入，但为了使其具有声门气流脉冲的实际波形，在它与道模之间还加了一个声门脉冲模型滤波器。随机噪声发生器就作为清音的激励源。浊音/清音开关用于控制激励源进入声道，以便产生浊音或清音，这个开关信号也应由外界输入。声道模型用来

模拟声道对激励源的调制作用，以产生各种不同的发音，其结构可以预先设定，但参数却无法预先知道，因为不同语音的参数并不相同，所以需动态调整。最后的辐射模型则用于模仿嘴唇的辐射效应，与声门脉冲模型一样都可以认为是不变的，可以预先设定。所以对于这个模型，因为不同语音无法预先设定的参数有下面四种：浊音/清音开关信号，若为浊音则还需有基音频率，声道模型的参数，音源的幅度 Av 或 Au(用于控制声音音量的大小)。对于这些参数的提取，是从对原始语音信号的分析中得来的。如果直接对这些参数进行编码，这样的编码方法就称为参数编码。但单纯用这种参数编码的方法效果并不太好，首先模型本身就是一个近似，离实际的语音发生过程有一定的差距；另外，参数的选择和提取也是很困难的，很难说所提取的参数就正好反映了信号的特征，所以还有个参数的优化问题。

为了解决上述问题，同时也参考了波形编码的优点，提出了分析——合成这种混合编码方案。它的基本思想是分析和合成，即通过对原始语音信号的分析提取出参数，再马上用合成的方法用这组参数去合成语音，通过比较合成语音和原始语音的差异，适时调整这组参数，使合成语音和原始语音在主观上听起来差别最小，参数也就成了优化参数。这种根据语音的生成模型，对于表征声源和声道的有关特征参数进行分析和提取，再运用这些参数重新合成语音的过程，就称为语音信号的分析合成，利用分析合成的方法进行语音编码的系统就称为声码器(Vocoder)。这类语音的声码器系统主要用于低速通信，它们可以高效地利用信道带宽，例如在普通电话线上，利用调制解调器就

可传输2.4kb/s到4.8kb/s的话音数据。因而声码器在军事通讯和其他通讯领域都得到了广泛应用，如前面提到的NSA的保密通信，还有欧洲的数字移动通信标准GSM，北美的数字移动通信标准CTIA等都用到了声码器。

到此为止，我们花了较长的篇幅介绍了语音的各种压缩编码方案，这不仅是因为编码是语音处理的基础，后面的识别、合成等语音处理都要用到它，而且同时也是因为语音压缩本身也具有重大意义，可用于提高信道利用率，减少存储空间，提高通信的保密性和安全性等等。事实上以压缩编码为基础的应用也十分广泛，如数字录音电话、电子留言簿、发声字典、多媒体查询系统以及各种各样能说会唱的发声玩具等等，无不留下了语音压缩编码的足迹。

第四节　各种各样的语音处理技术

本章一开始我们就指出，在与机器交往时人们希望能直接对话，现在随着信息产业的迅速发展，这种要求变得越发迫切：计算机、办公自动化、国防、通讯、机器人等无不要求有一个直接对话的“人机通信系统”。而另一项更加雄心勃勃的计划是要实现两种不同语言之间的直接国际通信，即通过“语音识别——机器翻译——语音合成”将一种语言直接转换成另一种语言，日本

和欧洲都有一些很强的研究组织在开发该领域的产品。我们不打算介绍这方面的情况，只想指出不论是直接的人机对话还是两种语言的直接国际通信都涉及到了语音处理的两个重要领域——语音识别和语音合成，这两个领域在过去的十几年中已取得了长足的进展，目前一些实验系统已在运行。另一方面的问题是在实际的通信中语音信号常会不可避免得受到噪声污染，这既有来自周围环境的，也有来自通信介质本身的噪音。在繁华喧闹的街头打过电话的朋友可能对此有过切身感受。其实不仅是通信过程会受到噪声的影响，而且对于一般的语音信号处理过程比如语音识别也常受噪声之苦。如何有效地去除噪声，恢复尽可能干净的语音就是语音信号处理的另一重要领域——语音增强的任务。语音增强技术同其他语音处理技术一样借助于硬件技术和本身的发展在过去的十几年中也取得了很多成果，开始应用于通信、识别等领域。

由此可见，语音处理发展到现在，已经形成具有不同应用背景的多个领域，比如前面已经提到的编码、识别、合成、增强等。而事实上正像在编码部分看到的那样，即使每一个大的领域尚有许多种适用于不用应用场合的技术，以识别为例，其下包括的就还有关键词识别、孤立词和连续语音识别、说话人识别等等。语音技术作为多媒体信息技术的一个重要分支，不论在通信、国防、机器人、娱乐、家电等领域都有重要应用，作为信息高科技产业的一部分，必将对我们的生产、生活产生巨大影响。下面我们就分别看看各个方面的语音处理技术。

一、会说话的机器

对于人会说话，即使很早以前，也不会有人对此感到奇怪；但对于能学人话的鹦鹉，人们就感到稀奇了；至于会说话的机器，恐怕没有人会无动于衷了，当然这多半是很久以前的事了，比如中世纪的欧洲，那时人们就认为让一个非生物体说话就是赋予其思想和灵性，从而被看作是一种神秘莫测的魔法，多见于神话传说之中。至于现在，接受过现代科学熏陶的人们大概不会再大惊小怪了，最多出于好奇罢了，因为不用说在电影电视科幻作品中已对能说会道的机器人相当熟悉了，就是现实生活中，各种会“说话”的机器也是屡见不鲜，手上戴的报时表、汽车上按的报警器、公共汽车上的报站器等，无不说得字正腔圆，听多了自然就见怪不怪了。

但有时人的好奇心还是很宝贵的，它能促使人们去研究问题，发明新东西。这不，语音合成的历史几乎可以追溯到 17 世纪，那时好奇的人们已经开始在探索人说话的机理了，还研制了“会说话的机器”。而且法国人 Mical 还真做出了几个“会说话”的简单装置。Mical 在 1778 年做的说话装置可以说一个长句子，而在 1783 年他做的一个能对话的设备，据说能说下面一段对话（原话是法文，这里直接给出中文译文）：

“国王给欧洲带来了和平”

“和平荣耀地为国王加冕”

当然限于当进的技术水平，效果不会很好，但说明了早在那时候人们就已在努力探索这个问题了。

进入20世纪，有关语音产生的理论取得了巨大进展，而随着电子技术的突飞猛进，20世纪70年代以后，LSI（大规模集成电路）和VLSI（超大规模集成电路）获得迅速发展，这为语音合成技术的发展奠定了物质基础，而数字信号处理与实验语音学，现代音韵学的交叉发展，互相促进，更是推动了语音合成技术的长足进展。语音合成是实现人机对话的关键技术之一，但要真正赋予电脑语言功能，必须彻底了解语言是如何产生和感知的。

语言的产生和感知既涉及高级的心理生理活动，也涉及较低级的物理、化学过程，处理过程非常复杂。语言的一般产生和被感知的过程是这样的，当说话人要表达一个想法时，总要进行语义处理，然后根据语义查找大脑中记忆的词库，把要表达的语义转化为句子（根据一定的语法规则和句法规则），然后由大脑向人的发声系统发出指令，以声道为中心，协调肺、喉、唇、舌等各部分运动，把这一蕴含着欲表达的语义的句子转换成语音信号说出来；语言的感知过程正好相反，它把收到的语音进行听觉处理和音位解码，把语音仍恢复成抽象的句子，同样通过查找词库，提取出句中蕴含的语义，于是说话人要表达的意思就被听者领会到了。研究语音合成技术，就是要模仿人发音的心理——生理——物理过程，让机器也能产生清晰而自然的语音。语音合成器就是建立在语音产生模型上的，它模仿了人的发声过程。

人的语音产生的声学模型在声码器部分已经介绍了，这里再简单回顾一下，它由三部分组成：声源激励模型，声道模型和

唇端辐射模型。语音合成器就是按这个模型模拟了语音产生的三个过程:声源激励,声道共鸣(声道指口腔、鼻腔和咽腔等狭长管道的总和)和口鼻辐射,其中模拟声道共鸣特性的数字滤波器尤为关键。

语音合成技术按基本合成单元的不同可分为四类,分别是波形编码合成,参数式分析合成,规则合成和文——语转换等,技术难度也依次递增,下面分别简要介绍一下。

语音的波形编码合成方式以语句、短语、词或音节为基本合成单元,这是比较大的单元,一般先分别录音后直接压缩编码存储起来形成一个合成语音库;重放时,根据待输出的信息,在语音库中取出相应单元的波形数据,串接或编辑在一起,经解码还原出语音。这种方式由于语音库中存储的是较长的语句、短语,数码率较高,占用存储空间多,所以一般合成的词汇量有限,但因为直接由长语音单元合成,所以自然度较好,只是连接处稍感不适,多用于自动报时、报号、报站或报警等装置中。

参数式分析合成方式用的合成单元较小,一般用音节、半音节或音素为合成单元。当然语音库中储存的也是这些较小的合成单元的语音参数的编码;而输出语音的过程与波形合成相似,也是根据待合成的语音的信息,从语音库中取出相应的合成参数、编辑和连接在一起之后顺序送入语音合成器,合成器就在合成参数的控制下,再一帧一帧地重新还原成语音波形。显然因为合成单元较小且用参数编码,自然数码率比波形合成方式低得多,但系统结构同时更复杂,合成音质也差一些,因为它毕竟是用参数合成的合成语音,恢复的已不再是原始语音了

采用更小的合成单元如音素、双音素等来合成语音时，语音库中除需要存储合成单元的声学参数外，还要存储诸如如何由音素构成音节，再由音节组成词或句的各种规则，这就是语音的合成规则，这种合成法就称为规则合成法。当输入字母符号时，合成系统就利用这些规则自动地将它们转换成连续的语音。与分析合成相比，规则合成方式的语音库更小，但音质也更差一些，而且系统结构也更复杂。

最后介绍的文——语转换系统才是真正实现人机对话的最佳方法，它以文本串为输入，根据合成规则合成语言信号，这也是一种规则合成系统。合成时，系统的文本分析器先根据发音字典，将输入的文字串分解为词及读音符号，再依据语义、语音规则，为每一个词、每一个音节确定重音等级和语句结构及语调、停顿等，这样，文本串就变换成为代码串了，规则合成器就可以依此合成抑扬顿挫和不同语气的语句。事实上可以看出，文——语转换系统除了依赖于多种规则之外，还必须依赖于对文字内容的正确理解，这就涉及到自然语言理解问题，所以真正的文语转换系统实际上还是一个人工智能系统。目前较满意的文语转换系统尚未开发出来，但这毕竟代表着语音合成的一个最重要的方向，也是人机对话的关键，所以现在人们正从各个领域进行努力探索，希望能早日实现人们的梦想。

二、"善解人意"的机器

随着欧洲在19世纪发明英文打字机之后，英文文字处理步入了机械化时代，但我们古老的汉字却如一个孤独的老人，怎么也进不了这个时代。由于汉字本身构造的复杂性和数量之巨，虽经能工巧匠、文人名士的努力，但造出的中文打字机并不实用，或用起来远不及英文打字机方便，于是我们古老文字的处理一直停留在用笔书写这种我们祖先几千年前早就发明的方法。计算机的发展，使我们开始进入信息时代，文字的处理也步入自动化时代，我们古老的文字终于把握住了这个时机，挤进这趟高速的信息列车。汉字编码，汉字输入，汉字编辑，汉字排版都取得了巨大进展，为汉字处理的自动化扫清了障碍。但是汉字输入的阴影仍没最终消散，尽管目前已有上百种汉字输入方案，但好学易用效率又高的方法却寥若星辰，目前汉字输入已经成了汉字信息处理的一个瓶颈，人们期望着自然的口述输入方式的出现，而且，这个要求现在变得越来越迫切了。可喜的是，希望的曙光也越来越亮，主要是因为语音识别技术随着信号处理和电子技术的迅猛发展已有了长足进步，并且逐步取得了实质性进展，简单的语音识别装置已经走向了市场，语音打字机作为语音识别的一个应用，虽说离实用尚有距离，但简单的个人口授系统也已出现(如IBM的个人口授系统)。而且事实上，语音识别的应用领域是十分广泛的，并不仅仅限于人机对话，还可以在通

讯、数据库电话自动检索、语声命令等领域发挥巨大作用，可以减少人们的重复劳动，提高效率，减少费用。这一些都促进了语音识别技术的迅速发展，也是为什么70年代以来美国国防部两次拆巨资搞ARPA和DARPA计划的原因。

早期的语音识别系统大多是按照模板匹配的原理工作的。

整个识别系统可以分为两个部分实际上也就是两个阶段。第一个阶段是训练阶段或称学习阶段，用户将词汇表中的每一个词说上一遍，计算机对所有词的语音信号进行预处理，提取出每个词的特征参数作为模板存入模板库；在识别阶段，计算机将输入语音进行同样的预处理提取出特征参数，然后把这个特征参数依次与模板库中存放的每一个模板进行相似度比较，就将相似度最高的那个模板所代表的词作为输出、完成输入语音的识别过程。由此可见，其基本原理和过程同汉字识别等图像识别过程有相似之处，毕竟它们都属于计算机模式识别的范畴，基本原理都是模式识别理论，只是处理的具体信号不同识别内容不同而已。在上面的介绍中，提到了模板库，或者更通俗地说就是词汇表或称词典，要指出的是，词汇表中提到的词未必要是我们平时所理解的词语（如果用术语"词条"可能更好区分一些），这是因为词汇表中的词条可以是字、词、句等较大的识别单元，也可以是音节、半音节、音素等比较小的识别单元。具体采用什么样的词条作为识别单元，也是语音识别中很关键的一个问题，因为这关系到词汇表的大小和能识别的词汇量的大小，比如用音素等较小的词条作为识别单元，不仅词汇表小，而且能识别的词汇量也大，但因用小单元构成词、句时常变化很大，识别起来

比较困难；相反若用字、词、句等较大的识别单元，则识别起来就比较容易，但能识别的词汇量就比较小了。如果硬要用大的识别单元来识别大词汇量的话，那么词汇表中词条的数目将会异常地大，因为所有与该词有关的短语、句子都得收集进去。

另外在语音识别系统中，说话人和说话方式也很重要。首先因为不同的人的讲话方式千差万别，即使是对同一个词的读音，亦会因方言区及个人文化背景、个人心理生理特征的不同而不同，但对于特定的某个人这些变化就小多了。所以对于针对许多人的非特定人语音识别系统做起来是比较难的，而针对特定人的识别系统就相对简单多了。到底什么时候采用什么识别系统这取决于实际应用场合，比如个人专用的语音打字机，就只要设计成特定人识别系统就可以了，而对于电话数据库检查系统，显然只能采用难度比较大的非特定人识别系统了。其次对于语音识别、说话人的讲述方式也很重要。被识别语音按述说方式可分为孤立词、连接词和连续语音，孤立词识别方式是指说话人每次只说一个词，一个词组成一条命令让识别系统识别。而连接词识别一般特指十个数字(0～9)连接而成的多位数字的识别，当然有时还可加上少量的操作指令，这比起独立词识别来要困难些。而困难最大的莫过于连续语音识别。连续语音识别是指说话人以日常自然的速度和方式讲话要求计算机予以识别，困难在于连续语音的各个发音单位往往是连成一片且上下文相互强烈影响，使其中的各个音素、音节和词之间的边界变得十分模糊，并不像孤立词识别系统那样可以明确地确定出来。

上面介绍了语音识别中较简单的模板匹配方法，和影响语

音识别的三个因素:词汇表大小,讲述方式和服务对象,其实这只是主要的三个因素,其他影响较大的还有语音的质量(噪声的强弱),识别器所需完成的任务(例如是关键字确认还是还需进一步了解语义信息)等。从上面的分析也可以看出,最理想的也是最困难的语音识别系统应是大词汇表、非特定人的连续语音识别,这时用简单的模板匹配法已不现实,需要使用目前作为主流的研究方法,即概率统计的方法。这种方法的基本思想是把从声学——语音层直至句法层,将全部语音的统计知识容纳在一个统一的框架之中。这种统计方法已被广泛地应用于语音识别领域,为 ATT 公司 Bell 实验室,IBM 公司,甚至美国国防部的 DARPA 计划都无一例外地采用了这一基本框架。另外随着人工神经网络、模糊理论等的发展,人们也正在研究如何用这些新的技术和理论来提高语言识别系统的质量,有人预测到 20 世纪末语音识别系统有望获得商业应用,也许到那时,我们的计算机会变得“善解人意”了。

最后咱们再谈谈语音识别的另一分支——说话人识别。说话人识别,顾名思义就是通过一个人的话音识别出一个人的身份来。它与一般语音识别基本原理相似,方法也差不多。两者区别在于,说话人识别并不在乎说话人说的内容是什么意思,而是希望从他的语音中提取出有关他本人的特征,换句话说,说话人识别是企图挖掘出包含在语音信号中的个性因素,而一般语音识别则是企图从不同人的语音信号中寻找共性因素,即说话人识别强调不同人之间的差别,而一般语音识别则力图对不同人说话差别加以归一化。

说话人识别的一个直接应用，聪明的读者也许马上就想到了，就保安领域，如机密场所的出入控制、通信线路的合法使用、机要设备的身份核查及电子银行转账业务的安全检查等，另外它在司法领域中可以帮助对嫌疑人员的查证以及战场侦听时对敌方人员的识别等等。目前说话人识别已取得阶段性进展，商用系统已经在若干领域中得到应用，但仍有不少问题有待解决，其中最关键的一个问题是究竟该用语音信号的哪些特征来表说话人才是有效而可靠的，这就涉及到了对人是如何通过听人讲话识别出人的这一过程的理解，可惜这个问题尚未最终解决，而且说话人识别系统的研制还涉及到许多学科和技术。当然这些问题既是困难也是挑战，它们的逐步解决必将推动这一领域的迅速发展。

三、去伪存真——语音增强技术

对于语音信号处理，我们前面已经讲得很多了，如语音编码、语音合成、语音识别等，但我们在进行上述处理时，都作了这样一个基本假设，即语音信号本身是“干净”的，所谓“干净”的意思是语音信号未曾被噪声污染。

正如人们生活的空间远非是干净的一样，语音信号的实际存在环境也远非是“干净”的，不用说周围的嘈杂声，就是通信介质本身也常常被多种多样的干扰、噪音所困扰，同样就像细菌会使我们人生病一样，环境噪声污染也会使许多语音信号处理系

统的性能急剧恶化，甚至根本无法工作。例如，目前已取得巨大进展的语音识别技术，正如前面所说的那样，正在步入实用阶段，但目前的识别系统大都是在安静的环境下工作的，一旦应用到实际环境中，将不可避免地受到环境噪声的污染，作为识别系统主要性能指标的识别率将大打折扣。另外低速率语音编码，尤其是参数编码，实际应用时也会遇到类似问题，因为参数的提取本身就是以语音为基础的。语音增强技术就是以此为背景迅速发展起来的，它的一个主要目标就是从带噪语音信号中提取尽可能纯净的原始语音。然而，由于干扰是随机的，从带噪语音中完全恢复纯净的原始语音几乎不可能，因为有的随机噪声本身就跟语音的某些部分十分相似，例如清音就表现出一定的随机噪声特性，别外还有其他语音的干扰也是较难去除的。在这种情况下，语音增强的目的主要有两个：一是改进语音质量，消除背景噪音，使听者乐于接受，不会感觉到疲劳；二是提高可懂度，但这两个目的往往是矛盾的。

从上面语音增强的两个目标可以看出，它不仅与语音信号的数字处理有关，而且也涉及到了人的听觉感知和语音学。另外，噪声来源众多，特性多异，很难找到一种通用的语音增强方法，能适用于多种噪声环境。所以只能针对不同噪声，采取不同的语音增强策略。下面就谈谈各种噪声类型以及我们去除它们的对策。

噪声来源于实际应用环境，常常变幻无穷，特性各异，按对信号的影响可分为加性噪声和非加性噪声两种，主要是加性噪声，对于非加性噪声有的可以通过变换转变为加性噪声，所以这

里主要介绍加性噪声。所谓加性噪声是指与语音信号叠加在一起的噪声，幅度上表现为两者之和。加性噪声一般有周期性噪声、脉冲噪声、宽带噪声和同声道其他语音干扰等。

周期性噪声和脉冲噪声的去除比较容易，因为周期性噪声在频域上表现为许多离散的线谱，可以用梳状滤源器加以滤除；而脉冲噪声表现为时域中突然出现的窄脉冲，所以去除脉冲噪声可以在时域进行。比如根据带噪语音信号幅度的平均值确定一个阈值，当信号幅度突然超过这一阈值时，可判断该时刻出现了脉冲噪声，对此进行适当衰减或平滑就可以了。

对于后面两种加性噪声的去除就没那么容易了。宽带噪声的来源很多，热噪声、气流噪声（如风声，呼吸声等）以及各种随机的噪声源等，模拟信号数字化时引入的误差（称为量化噪声）也是其中一种。由于宽带噪声与语音信号常在时域和频域上完全重叠，因而消除它最为困难。而同声道语音干扰是指有多人同时讲话时，他们之间的相互干扰。对于多人同时讲话，如果是在实际环境中，则因为人所具备的“双耳效应”完全可以把一种语音从众多的说话声中拾拣出来，但当多个语音叠合在一起，在单信道中传输时，双耳信号因合并而消失，人就无法区分开了，从而也就听不清某个人在讲啥了。如同人的指纹那样，每个人因发声系统生理构造上的差异，也有不同的“声纹”，这种声纹在于他们的语音生成模型参数的不同，藉此我们可以把我们所需要的某人的语音提取出来。

对于去除最难处理的宽带噪声，尽管在语音增强理论上尚未完全解决，但某些有效的增强算法已经提出来了，一般可分为

四类：噪声对消法，谐波增强算法，基于参数估计的再合成法和基于语音短时谱估计的增强算法。我们不打算一一介绍各类算法，因为它们涉及太多的专业知识，只想简单介绍一下双话筒噪声对消法，使读者对语音增强有一个感性认识。

双话筒噪声对消法的基本思路是用两个话筒，同时从现场采集带噪语音和噪音，进过适当处理之后，把两者相减，从而可以获得较干净的语音。这种方法对于强背景噪声环境，如纺织厂、汽车飞机上、坦克内等等，增强效果比较好。这种噪声对消技术的关键在于如何从环境中采集到足够"逼真"的噪声，所谓"逼真"是指与带噪语音中的噪声分量尽可能相似。

上面简单介绍了针对不同加性噪声的语音增强方法，语音增强同其他语音信号处理技术一样，应用领域也十分广泛，不仅可作为其他语音信号处理的预处理，而且还可以应用于许多其他场合，如去除移动通信中的噪声干扰，室内电话会议的交混回响问题，有历史价值的唱片、旧录音带的失真和噪声去除，有语言障碍残疾人的语音失真的校正等等都可以找到语音增强的用武之地。

第五节　计算机声音媒体的广泛应用

正像我们在本章开始时就指出的那样，随着计算机声音处

理技术的迅速发展，同时伴随着我们步入信息时代，计算机声音媒体的应用越来越广泛，它不仅应用于传统的通信、人机对话、影视娱乐等领域，而且还扩展到银行金融系统、证券税务系统、音像出版业、计算机音乐界、国防事业、工业控制等经济、政府、文化、国防、工业等部门和领域。更值得注意的是，它还逐渐进入我们的生活，为我们的生活提供了诸多方便，使我们的生活变得更加有声有色。

通讯是语音的传统应用领域，因为目前的广播、电视、电话等通讯系统无不全部或部分以此为信息媒体。但近几年来，随着通讯方式的不断改进，尤其是随着计算机与通讯技术的紧密结合，更是涌现出了许许多多更为先进的通信方式，使语声技术的应用前景更加看好，因为语声通信毕竟是人们最方便最直接的手段。比如计算机网络技术的发展，使数据通信变得十分便捷，电子邮件（E—mail）的出现，使邮件通信变得十分方便。传统信函少则几天，多则十天半个月才能到达收件人手中，而电子邮件视网络的拥挤程度一般只需用几秒钟到几十秒钟，最多也不过几分钟就可到达收信人手中。而且令人惊奇的是，这种电子邮件还能传寄语音信件，即语音电子邮件。语音电子邮件的核心就是语音的编/解码技术。发信方只要把欲传递的一般话音录下来，由计算机进行压缩编码之后形成一个数据文件——电子邮件，然后直接从网络上发送出去；收信人收到邮件之后，由计算机用同样的编解码方法进行解压缩恢复成原来的话音，于是远方亲人的殷殷关切之声即刻在耳边响起，岂不令人心襟动荡。现在，举世瞩目的互联网络（Internet）已提供了语音电子

邮件服务，你可以把你的话音送往世界几乎每一个地方，而不用担心高额的长途电话费，因为互联网络比长途电话便宜得多了。而且据1996年年初的消息，已经有软件供应商在开始提供可供在互联网上打电话的软件，这种软件的核心之一也是语音的编/解码技术，因为互联网上调制解调器的速度只有14.4kbps，语音不经压缩显然是不行的。

即使在传统电话网中，也已开通了多项语音服务内容，如电话费的自动查询系统，电话语音信箱，电话语音留言等等。从1995年开始推出的用市话线打长途的200卡，也部分应用了语音合成技术，用来向用户报告账户上还剩多少钱。而经济、政府等部门运用电话处理日常业务，又推出了许多应用语音技术的新的业务经营方式，如电话银行电话查询系统、信用卡电话授权查询系统、证券电话委托交易系统、货运到站查询系统、税金催交系统等等，这些新型服务方式不仅大大方便了顾客，树立了企业政府的良好形象，还大大提高了工作效率，创造了社会经济效益。以信用卡电话授权系统为例，在应用该系统之前，当要验证时，管理员根据查证者打来的电话，手工输入有关信息，再把查询结果用电话告知查证者。如果应用了电话授权查询系统，管理员就省去充当中间角色的任务，而由系统自动识别出查证者用电话送来的语音信息，同时又把查询结果用语音的方式告诉对方，实现整个查询过程的自动化。考虑到在发达国家，信用卡的授权查询是一项经常的业务，应用这套电话授权查询系统的社会经济效益是不言而喻的。

随着科学技术的发展，现代化的战争也越来越多地依赖于

现代科技，出现了形形色色的使用各种现代化装备的部队，于是如何才能有效地协调好各种军事力量，有效地打击敌人成了一个很关键的问题。作战指挥自动化应运而生，而作为指挥自动化的"脉搏"的通信也就日益显示出其在战争中的特殊地位，这也为语音技术的发展提供了契机，因为军事通信中本身就要用到大量的语音技术。比如保密通信中要用到语音加解密技术，信道使用中的安全检查要用到语音识别技术，战场强噪声环境中通信要用到语音增强技术等等，更不用说作战指挥系统中大量语音情报的收集、整理、发送了。这也许正是美国国防部不惜以巨资研究语音技术的重要原因之一。

除语音之外，其他声音媒体的应用也大有市场。君不见，各种各样光盘存储器的出现，在音像出版业中又掀起了一场"数字化革命"；计算机合成音乐以其独到的魅力，不仅令众多的"音响发烧友"为之倾倒，而且还极大地拓宽了音乐家的创作天地；电子游戏业的蓬勃发展，也为计算机声响尽现迷人风采提供了广阔的天空，各种似真似幻的人造音响效果，使你在使尽浑身解数与"街头勇士"(一种游戏)们奋力拼杀的同时，还可犹如身临其境地听到各种逼真的模拟现实格斗环境的声响——呼呼的抡拳声，叮叮当当的武器撞击声，气喘吁吁的呼吸声等，甚至还能听到有人从背后向你袭击时擦破空气的轻微风声。当然面对如此诱人的游戏，朋友们切不可因此沉迷其中，耽误了学习和休息。

总之，作为多媒体技术之一的声音媒体技术不论在今天还是明天，不论在生产领域还是日常生活，都将以其得天独厚的优势继续发挥其应有的作用，不断地造福于人类。

第五章　五花八门的多媒体应用系统

1981年，当IBM公司出乎意料地首次发布IBM PC机时，恐怕谁也不曾料到，短短十几年，PC机的销售量竟会奇迹般地达到一亿多台；当然更没想到的是PC机竟会如此大地影响到整个世界。不用说PC机对国民经济、对科学技术的推动作用，单就其对人们工作方式、思维习惯的改变来看，无疑也具有革命性的影响。如果说PC机的出现还只是改变了人们的工作方式的话，那么多媒体技术的发展，则将进一步改变人们整个的生活方式。所以说20世纪80年代还是微机时代的话，那么20世纪90年代就是多媒体时代。

事实也是这样，随着多媒体越来越多地成为工业界、新闻界以及日常人们谈论的话题的同时，多媒体也在不知不觉之中把触角伸向了我们生活的每一个角落。比如以光盘为代表的电子出版物犹如雨后春笋纷纷上市，大有与传统出版业一争天下之

势，而多媒体教学系统也在一片啧啧声中“粉墨登场”，踏上了神圣的讲台；各种交互式多媒体系统亦施展开各种“交际手段”，甚至不惜“软（件）硬（件）兼使”，双管齐下，把好奇的观众轻松地拉到自己交互式的屏幕面前；而多媒体视频会议系统更是有声有色，一会儿是会场瞻望，一会儿是发言人特写，一会儿又把文件摊开在“桌面”上让大家讨论，“玩”得人眼花缭乱，以至于与会者忘了自己究竟身在何方了……诸如此类，不胜枚举，多媒体的应用空间几乎是无限的。

目前，多媒体已广泛应用于辅助教学（如交互式视听教材、模型、过程演示等），电子出版物（如光盘百科全书、多媒体词典、电子报纸、文献资料、音乐美术库等），多媒体通信（如各种电子邮件、可视电话、多媒体会议系统等），多媒体地理信息系统（如铁路管理信息系统、电信综合管理系统、电网综合信息系统等），多媒体监控系统（如各种安全检查系统、夜视监控系统、现场监控系统、指挥自动化系统等），管理信息系统（如人事档案管理、病历管理等）等领域，另外在日常生活中多媒体还广泛应用到电视、广告、演示、旅游、娱乐等我们早已熟悉的领域。由此可见，多媒体技术确实已经或正在在我们的许多领域中获得广泛的应用，并正慢慢地向我们的日常生活走来，在不知不觉之中改变着我们的工作方式和生活方式，使我们的生活变得更加绚丽多姿。既然通向多媒体时代的大门已在向我们打开了，那么为何不让我们走进去，看看多媒体到底为我们做些什么呢？

第一节　多媒体辅助教学

眼睛是心灵的窗口，我们常以此来比喻眼睛所独具的魅力，事实上，眼睛不仅是通向心灵的窗口，也是外界信息进入我们大脑的窗口，因为用眼睛看是人类最有效、最重要的获取外界信息的手段。也许大家听说过这样一个事实，人类所接受的外界信息中约有80%来自视觉。另外现代教学实验也证明：学生如果仅靠听觉获取知识，3小时后能保持70%，而72小时之后只能保持10%；而如果只靠视觉的话，3小时后能保持72%，72小时后仅能保持20%；但如果视听结合起来，3小时后保持85%，72小时后还能保持65%。由此可见，采用多媒体教学，将图像、文本、声音、视频及音频等结合起来，提供一个图、文、声并茂的交互式教学系统，使学生由被动接受变为主动学习，那么其教学效果肯定是非同寻常的。

事实也正是如此，日本中等学校的英语教学，从1993年4月起就开始采用多媒体英语教学光盘。这种英语教学光盘上的内容包含有文、图、声等媒体信息。上课时老师按教学盘讲解，课后学生在校或在家里用这种光盘进行学习，从单词发音、词义解释、遣词造句到整篇课文的诵读均是图、文、声并茂，并可反复再现。不懂的地方，只需向多媒体系统提问，它就会热情地、耐

心地向你详细解答，不懂的地方可以多问几次而不用担心系统会不耐烦，更不会嘲笑你、责备你。当然你也不能过于放松自己，对于你的学习能力、学习态度以及你的想象力，系统都会一一记下，俨然就是一个作风严谨的好老师。

这种多媒体教学系统的最大好处在于其图文声并茂和具有交互能力，能充分调动你的五官，犹如身临其境般地让你去感受、去体验，甚至可按你自己的水平和能力选择学习，因材施教，充分激发你的学习自信心，循序渐进，效果自然不同凡响。目前国内开发的《动态英语》也是这种具有交互能力的多媒体 CD—ROM 节目，其他还有双语资讯公司的《桌上英语学校》系列多媒体外语教学软件等。

多媒体教学节目不仅可应用于外语教学，还可应用到其他学科教育，如我们在动画部分介绍的地理教学系统，还有历史教学、生物教学、音乐视唱等。比如生物多媒体教学节目，不仅可以用文字向你讲解各种动、植物的知识，还可以使你具体看到动植物的形态，听到动物的叫声，如鸟鸣声、狮吼声、虎啸声、蛙鸣声、狼嗥声等等，甚至还可以看到根据化石恢复出来的古生物的形状及模拟的远古生态环境。又比如历史多媒体教学系统，使你免除了枯燥地按年代记忆史实，相反可以直观地犹如亲身经历似的观看历史发展过程。比如可以看到陈胜、吴广揭竿而起的悲壮场面，听到二次大战隆隆的炮火声，呼啸的飞机声，还可以看到人类首次登上月球，听到阿姆斯特朗“我只是跨出了一小步而人类却跨出了一大步”的壮严宣告，甚至亲自聆听肯尼迪总统感人肺腑的名言：“Ask not your country can do for you, ask

what you can do for your country。"

同样，孩子们也不用再守在电视机前被动地收看《太空堡垒》或沉迷于各种与学习无益的电子游戏中。例如，有种 35 分钟的《太空旅行》多媒体 CD—ROM 游戏节目，它可让您在漫游银河系的同时，扮演一位已陷入黑洞的飞船的船长，让您在三维空间中探险，使孩子们在玩游戏的同时，接受现代科技的熏陶，做到了寓教于乐。

多媒体引入教育，不仅改变了学生的学习方法，也改变了教师的教学方式，改变了教材的编写方法，从而改变了现代的教学模式，使由教师为中心的被动教学模式发展到以学生为中心的主动教学模式。

多媒体应用与教学相近的一个领域就是培训，通过多媒体光盘教学系统营造出一个受训环境，甚至用多媒体网络营造一个实时多媒体培训环境，将使受训人员免除奔波之苦，使公司免除大笔的培训经费，而且使培训变得更为容易，在大大节约时间、金钱的同时，也提高了培训效果。

第二节　多媒体电子出版业

在北宋庆历年间（公元 1041～1048 年），毕昇发明了活字印刷，这个发明对于推动人类文明和进步曾起到了巨大的作用，为

我们留下了一个珍贵的思想文化宝库。现在不仅汉字可以进入计算机，其他像图像、声音等都可以以数字的方式进入计算机，这使我们的出版业又经受了一场划时代的巨大变革，开始告别铅与火的时代，迎来了光与电的世纪。

目前的出版物主要有以下几类：1. 传统的纸面出版物；2. 以缩微胶片、录音带、录像带等影像制品为代表的非纸面出版物；3. 以电、磁、光等为信息载体的数字形式存储信息的电子出版物。对于前面两种出版物，大家已经比较熟悉了，对于第三者也许就不那么熟了。那么什么是电子出版物呢？

电子出版物就是以图、文、声等多种媒体形式表现的并且由计算机及其网络对这些各种形式的信息以内在的统一的方式进行存储、传送、处理以及再利用的出版物。电子出版物与传统出版物相比具有如下特点：首先电子出版物往往以光盘形式发行，存储量巨大；一张光盘存储几百本长篇小说并不是件难事，其次电子出版物运输和携带方便，体积小，重量轻，并可长期保存；第三可以及时传播，经由计算机网络可即刻发往国内外各地；最后节目组织形式灵活，采用多媒体技术具有信息漫游功能，即具有很强的交互阅读功能。

目前电子出版物种类繁多，几乎包括了现今所有的传统出版物的形式，不过，为方便起见，我们把它们分为下面几类：电子图书、电子期刊、电子新闻报纸、电子游戏软件、电子文献及检索工具、电子音像制品、电子公文等。下面就分别具体介绍这几种形式的电子出版物。

一、电子图书

目前电子图书一般以CD—ROM光盘的形式发行，品种多样，内容十分丰富，从人文社科到自然哲学都有产品问世。

多媒体光盘图书与一般图书相比，不仅图、文、声兼具，可以具有极强的交互能力，而且其内容组织往往不同于一般的图书，也不同于影视表现，某种程度上可以说是兼有图书和影视两者的优点，即在内容组织上采用了超文本技术。

大家知道传统的图书是一页一页翻着看的，而影视表现中则常采用"蒙太奇(Montage)"手法，即影视表现中更注重于通过流动的语言来表达主题，内容的组织是动态的，可以从一个画面跳到另一个相关联的画面。超文本数据管理技术就采用了这种影视表现手段，更确切地说是模仿了人类的联想记忆功能。采用超文本技术管理多媒体信息就形成了超媒体技术。超媒体图书一般看起来较符合人的思维习惯，令人有身临其境的感受。比如，美国麻省理工学院研制的《白杨城影片地图》就是这样一种超媒体系统，它用来模拟旅行，用户可以利用它在计算机屏幕上模拟"驾车"游览白杨城。该地图系统装有一组包括白杨城所有街道照片的视频盘片，它们是通过按装在卡车上的四个摄影机，沿街每隔3米各取一个镜头而获得的，而且每一照片都与其他一些相片相联系。人们可以连续向前、倒转，或者左右移动地观看。用户使用该系统感觉就像驾车通过该城市一样，可以在

任一十字路口转弯，甚至可以走进建筑物里面去看看。显然要达到这样的效果，信息的组织肯定不能像一般书那样是一页一页的，而必须具有“联想”功能的，能从一种场景自然过渡到另一场景，比如在建筑物前，接下来可能的“联想”就应是走进去并显示出里面的场景。超媒体实际上作为一种技术，也可以应用在其他媒体系统中，比如辅助教学、辅助培训、产品概览、旅游指南，甚至虚拟现实等。不过目前像《白杨城影片地图》这样的超媒体系统还不多见，比较实用的是各种超媒体词典、手册、文献资料等。比如词典中相关词条如近义词、反义词等的自然“联想”；或者比如看到词条“地震”，除给出文字解释外，还可以放一段地震形成过程的动画，甚至是现场拍摄的地震场面，使读者不仅对地震的词义有了深刻了解，而且还对地震的成因及危害有了一些更深刻的感性认识。

现在各种形式的电子图书已经很多了。比如，《未来图书馆》一张光盘就存在450多部文学哲学名著，内容包括从亚里士多德到康德、从荷马到伏尔泰的任何篇章；而另一部著名的多媒体光盘作品是《爵士乐的历史》一书，它不仅详细地介绍了爵士乐发展的过程，而且还同时介绍了爵士乐发展过程中有影响的乐队和乐手，并在介绍他们各自的风格时，还引入了大量的演出录音和录像，极大地提高了作品的感染力。目前国内也推出了多种光盘图书。比如，杭州矽谷光盘有限公司最近推出了一批电子图书，其中，《实用百科》包括动物篇、植物篇、科技篇、气象篇、天文篇、地理篇、人体生理篇、生活常识篇等；而《成语故事》则收集了1000多个成语故事，并配有优美的音乐和生动逼真的

插图。再比如，北京金盘电子有限公司也已开发了英语教学光盘《动态英语》《邮票上的中国——历史和文化》等电子光盘图书。

二、电子期刊

对于杂志，相信大家都不会陌生，我们知道现在的杂志一般都是通过邮局投递到我们手中的，这往往要不短的时间，而且邮资也不菲；另外对于这种传统的纸面杂志它只能提供静态图片和文字，对于活动图像和声音则无能为力只好割爱了。这两条缺憾对于一般的杂志也许算不了什么，但对于新闻信息类等时效性较强的杂志第一个问题就比较突出了，而对于影像、音乐杂志第二个问题的存在简直是致命一击，使其失去了原本应有的魅力。多媒体电子期刊的出现，才算弥补了这两个缺憾，因为图、文、声并茂正是它的特点所在，而且容量巨大，内容异常丰富，如果再经网络发行的话，那么我们在家中就可以在刊物出版之日马上挑选自己想看的内容阅读，免除了因投递而造成的时延。

目前电子期刊的发行方式有两种，即光盘版和网路版。光盘版就是以大容量的CD—ROM为存储介质装入期刊内容，而网络版则是直接把本期内容放在某个服务器上，合法用户可以通过网络进入该服务器，阅读杂志上自己感兴趣的内容。目前这两种形式的电子出版物已经很多，下面就举几个例子介绍

一下。

美国《新闻周刊》是一种光盘杂志。它每季度定期发行可供多媒体用户使用的CD—ROM盘片，每张盘上除存有过去三个月中每期杂志上的文章外，还有二至三篇原杂志上所没有的专题报道文章。国内第一种以光盘为载体的电子期刊《电脑视野》，也于去年(1995年)下半年由北京大恒图像视觉有限公司出版发行。该光盘杂志融合了图、文、声等多种媒体形式，具有传统报刊、杂志和广播电视的效果，对于市场上和技术上的热门话题，光盘杂志也进行专门分析，并配以相应的演示和应用软件，使您可以边看边研究。而台湾轩丞公司于1993年3月创刊的多媒体杂志《WINmagazine 佐品杂志》则更有特色。该杂志于1993年8月份开始刊出人物访问记，可使那些被访问者当时的形象和原声，栩栩如生地重现在读者面前，令读者看起来饶有兴趣，并感觉比电视还要真切。

对于网络版电子期刊，限于国内网络通信的水平，尚未听说有杂志发行，但在美国由于网络联机服务相对比较普及，现在已有不少这样的杂志，《时代周刊》就是其中一例。《时代周刊》于1993年9月开始发行电子版，通过美国联机服务公司将杂志上的所有文章传递给用户。它在互联网上设有两个网络地址，供合法用户进入其中任一地址阅读经常在更新的新闻报道及分析。《美国新闻与世界报道》电子版是另一例电子出版物，它也包括了印刷版本的主要文章、图片照片资料等。电子版用户通过电脑信息服务公司的联机服务，在印刷版本送到订户手中之前便可阅读到杂志上的文章和图片，享受到先睹为快的乐趣。

三、电子新闻报纸

随着20世纪90年多媒体技术的蓬勃发展，其应用领域迅速扩大，作为我们社会信息主要来源之一的新闻报纸业，自然首当其冲，成了多媒体小试“牛刀”的好地方。大家知道，新闻业作为一种特殊的信息处理行业，其对新闻信息的采集、处理、传递、广播、出版等活动无疑与一般的信息处理过程有着很多相似之处，那么多媒体技术自然在新闻作业中也应大有用武之地。事实也真是如此，多媒体技术在新闻采编制作过程中的应用已导致了多媒体新闻创作平台和多媒体无纸采编网络的问世，而多媒体在新闻广播、新闻出版业中的应用则直接造就了一种全新的新闻传播模式——多媒体新闻。对于前者因涉及较多的新闻技术、网络技术、计算机技术等专业技术领域，我们就不介绍了，下面主要介绍一下跟我们日常生活关系比较大的多媒体新闻。

多媒体新闻在某种程度上可以说是传统报纸新闻、广播新闻、电视新闻的集大成者，具有图、文、声并茂的特点，并且还有独到的优点：更好的时效性、可选择性、可重播性、传播方式的不同性等。目前出现的多媒体电子报纸正是这一全新的新闻模式的最好例子。

电子报纸或称报纸电子版通过计算机网络传递，用户利用电话线即可将本地电脑与报社发行部的服务器相连通，通过调制解调器就能收看到当日的报纸，这不仅使用户可以看到图文

并茂的报纸，还可以享受到先睹为快的乐趣。据统计，目前通过网络（Internet）阅读的报纸已有900多种，估计到今年（1996年）底将达到2000种，发展速度之快令人咋舌！世界上许多有名的报纸已纷纷先后发行了电子版报纸，比如美国的《华盛顿邮报》《纽约时报》《波士顿环球报》、西班牙的《阿贝赛报》等等。我国的人民日报也从1993年起按年度发行《人民日报全文数据库》光盘版，而且也提供了灵活实用的检索查询功能。

四、电子游戏软件

对于电子游戏，现在的年轻人不会太陌生了，有的人甚至还非常熟悉，说起来如数家珍。不过我们要讲的多媒体游戏不同于一般的街头游戏，那是专门的游戏机，这里谈的是计算机上玩的软件游戏。

事实上多媒体技术的发展有一部分也得益于游戏软件的发展，因为游戏往往追求生动有趣，常大量运用图像、计算机动画及音响效果，甚至在西方有一种流行的看法，认为最好的多媒体软件设计是通过计算机游戏体现出来的。

早期的游戏限于显示器的分辨率和计算机的速度，图像和声音都比较简单，但有些游戏至今仍受人喜爱，比如“挖金子”、“吃虫子”等。后来，好的音响效果被用于游戏，装有声效卡的计算机在玩游戏时可以获得逼真的音响效果，比如比较著名的有“三国演义”、“三维狼”等。而当多媒体计算机的标准公布之后，

游戏软件的开发者便把注意力转移到多媒体游戏上。下面介绍两种典型的游戏。

"国际象棋"游戏是一种下棋游戏,其模子造型是三维的,如皇帝、皇后、士兵等均栩栩如生。当移动一步棋子时,计算机就会放出人走动时的脚步声或特定的音乐。当一方吃掉另一方棋子时,就会伴有双方格斗的画面,刀光剑影,人喊马嘶,中剑者发出的惨叫,无不令人惊心动魄。整个下棋游戏生动活泼,妙趣横生。1996 年 3 月下旬,中国棋院为推广桥牌、国际象棋也推出了这种类似的电子出版物。该出版物由国际象棋大师和桥牌大师主讲,从最基础的棋牌知识到大师们的精彩对局都有,对局或对弈就采用了上面这种生动有趣的方式,不仅造型别致,而且还有很好的动画效果,你甚至还可以与计算机一决高下。

"英语看图识字"则是一种教育游戏。游戏中有许多生活中常见物体的场景,孩子们可通过与计算机"对话"学习相应的单词。例如屏幕上出现 9 个球,当孩子将光标指向绿球时,计算机就会发出标准的英语读音"Green Ball。"孩子学完之后,计算机就会把球的秩序打乱,然后问"Where is green ball?",让孩子自己去找。找错了可重试,找对了就把它拿走,再找余下的球。总之,使孩子们在轻松的游戏环境中不知不觉学会了单词和发音,真正是寓教于乐。

目前游戏软件很多,多以 CD—ROM 光盘的形式发行,往往一张盘中包含有许多游戏软件,不过现在的游戏软件也越做越复杂,效果也越来越好,软件的数据量也越来越大,一张盘能装的游戏也越来越少。比如"福尔摩斯探案"是一张风靡世界的

多媒体光盘。在这个游戏中，你可以充当其中的一个角色，同福尔摩斯一起破案。而且这张光盘不仅有图像、声音和动画，还包含有90分钟的活动视频画面。我国大陆游戏软件的起步较晚，生产尚不太多，但近年来发展也很快；而在宝岛台湾游戏业则是一片繁荣景象，比如有名的“三国志”、“三国演义”、“七侠五义”、“笑傲江湖”等均出自台湾游戏软件制作者之手。

五、电子文献及其他

谈起文献，也许大家马上会想起那厚厚的大本年鉴、一摞几尺高的百科全书、蝇头小字的科学文摘及其检索工具等等，这些资料的一个突出特点是数据量极其庞大，若无自动工具查找比较困难。

光盘出版物及其检索工具的出现，为这些文献资料的管理和使用提供了方便。比如美国《中央情报局年鉴》，您可以从这个数据库中方便地得到248个国家和地区中的任何一个国家和地区的详细资料，比如地理、气候、人口、政治、经济等方面的信息；而利用《美国鸟类》百科全书光盘，您可以在看到对各种鸟类的文字描述的同时，还能看到1840幅精美插图，听到115种鸟的叫声，这是传统的纸面出版物所无法做到的。事实上如今图书馆收藏的不仅仅是书、报刊杂志、手册等一类纸面出版物了，CD—ROM光盘等电子出版物也已成了一类主要的收藏对象。光盘图书的另一个好处是检索方便，通过检索工具提供的友好

用户接口，用户只需用鼠标、键盘、触摸屏甚至语音与机器交互，多媒体系统将为您提供完整的服务；而传统的卡片检索却要一盒一张地翻阅卡片，还需要用笔作记录，不仅繁琐，而且往往低效。比如科研院校经常使用的《工程索引》、《科学文摘》等文献检索工具，对于原先的纸面检索工具，查找一个文献至少得几分钟，有时甚至十几分钟，几十分钟；如果用现在的光盘检索工具，即使关键词确定的范围过大，多查了几次，也不过几秒钟或数分钟而已，检索效率的高低对比是显而易见的，而且后者不需要人的干预，免去手工查找之苦。

电子出版物的其他形式还很多，比如，像 V—CD 盘、DVD 盘那样的音像制品，CD—I 盘、DVI 盘那样的交互节目，还有电子公文、电子相册、电子画报等等，不仅品种多样，内容多异，而且发展神速，前途无量，在出版业中掀起了一股电子浪潮；并进而也影响着人们的工作方式和日常生活，是多媒体产业中一颗瞩目的新星！

第三节　多媒体通信系统

如果说 20 世纪 80 年代初由 IBM 公司首创的 PC 机使计算机个人化的话，那么 20 世纪 90 年代多媒体技术和网络通信技术的发展使计算机系统走向“分布的集成化”，即网络把分散的

系统联在了一起。由此而产生的各种多媒体通信系统将大大方便我们的日常工作和生活，并改变人们平常的工作活动方式。比如小组合作以及协同工作的工作方式就是在这背景下提出来的新的项目开发方式，它使不同地区不同行业甚至不同国籍的科研、管理人员能够在多媒体网上联合攻关，协同工作开发一个新的项目或合作开展工作。另外多媒体通信技术的发展也将使我们的个人通信方式得到极大改进，由普通电话向着可视电话、可视通信迈进。下面我们就来看看目前的三种多媒体通信系统：多媒体会议系统、视频会议系统和个人“桌面”可视通信系统。

一、多媒体会议系统

多媒体会议系统是在计算机技术、多媒体技术和通信技术的基础上发展起来的一种多媒体实时通信系统。多媒体会议系统的优点是不难发现的，它可以避免传统的面对面坐在一起开会的许多限制（比如长途旅行、来回奔波等），可以节省时间、精力和资源，于是有人把这种会议称为“绿色会议”；另外因为无需离开自己的工作环境和同事，与会者还可以在开会的同时兼顾一些日常工作；会议系统缩短了远地合作者之间的距离，加速了响应和决策的速度。也许大家会问，以前不是也有电视会议嘛，它也有这些优点呀？问题在于电视会议系统虽可解决异地开会的问题，但它只能在电话局的特别房间内进行，并需租用昂贵的

通信线路，而且它的信息传输方式也不灵活，一般传送数据文件比较困难，更谈不上对信息的即时加工、共同决策了。

目前各种实验性的、应用性的多媒体会议系统已不少，如局域网上的有：美国 AT&T 公司的 Raport 多媒体会议系统，美国 Xerox Polo Acto 研究中心的 Ether phone 多媒体会议系统，日本 NEC 的 MERMAID 多媒体会议系统等，广域网上的比较少，著名的有美国 DARPA 远程会议系统和欧洲的多媒体综合会议系统 MICE(Multimedia Integrated Conferencing for European Researchers)。

各种多媒体会议系统一般都有自己的硬件和软件结构，但可以发现系统基本上都由两部分组成：会议服务器和客户机。

会议服务器用于控制会议的建立、进展和终止，以及管理与会者的加入与离开、发言请求和传送等，是整个会议系统的控制核心。另外在某些系统中还有一些专门服务器，用以管理多媒体信息及提供通信服务。客户机则是与会者参加会议的“窗口”，代表了虚拟的会议房间，用户请求参加会议或离开会议以及发言申请等均有客户机的窗口界面管理，由它再把用户的请求传给服务器，服务器对此做出的响应也由它显示出来。

多媒体会议系统为用户提供两种主要服务：第一是用户之间的实时通信能力，在不同机器和网络之间共享可处理的多媒体信息；第二用户访问一个或多个服务器，即分布在不同机器和网络点的计算机资源。召开会议时，包括发起者在内，至少应有两个与会者参加；而每个用户都可以发起一个会议，这只要在用户工作站某个窗口输入邀请来参加会议的用户姓名即可。当然

被邀请的用户可以接受邀请或拒绝。

多媒体会议系统的应用除了召开一般意义上的会议之外，还可应用于各种讨论会、碰头会，可以支持办公自动化、远程教育、工程设计等。比如英国伦敦大学 UCL 的 CAR 项目就是用于汽车工业的多媒体会议管理和支持系统，其目标是支持分布式合作设计汽车，以缩短汽车的开发周期。据统计表明，办公室人员一般几乎有一半时间花在会议上，所以开发支持小组决策、协同工作的多媒体会议系统是很有意义的。

二、视频会议系统

我们在第三章图像媒体的视频部分已初步介绍了视频会议系统，指出它是解决通讯有效性的一种很好的途径，下面来具体看看其应用现状和前景。

视频会议系统是视频图像、语音等多媒体信息数字处理与数字通信技术相结合的产物，已广泛应用于军事领域、政府机构、贸易公司、医疗系统以及个人家庭等。

从前面几章的讨论中，大家也许已经意识到，视频会议系统的难点在于传输多媒体数据所需的很高的传输速率上，尤其音频、视频等数据量大，又要实时传输的多媒体数据更要求极高的传输速率。所以早期的视频会议系统由于使用的专用芯片昂贵，同时又要占用很宽的通信道带宽，使其推广使用受到很大限制。后来随着数字压缩技术的发展，同时高速率的通信网也逐

步建起来之后，才使视频会议系统真正步入了实用阶段。正如前面所介绍的那样，美国利用ISDN(综合业务数字网，它能较好地完成多媒体数据的传送任务)、数字交换网和专用数字网连接视频会议系统，并已开展了出租业务，只需很少费用即可在相距很远的两地“面对面”谈话。

一般视频会议系统的组成大致可分为三大部分：

1. 多媒体信息处理计算机及其输入输出设备，这是各个会议点都要配置的，这部分主要包括：

(1)档次较高、配置较好的微机。

(2)实时动态视频、语音信息的输入、压缩、还原及输出的专用卡。它的基本要求是：提供动态视频，语音信息的输入/输出，压缩/还原；支持广播级的图像输出并具备画中画功能；提供与数字通信网设备的标准接口及控制输入/输出设备的专用接口。

(3)输入/输出设备，包括显示器、音响系统、摄像机、录像机等。

(2)多点控制器。它的功能有点像电话局的交换机，只安装在主控台，用于各个会议点之间的实时多媒体信息交换，并对会议点的个数，数字通信的带宽，会议的控制模式(语音激活模式，主席控制模式还是讲课模式)、系统起动和故障诊断参数等进行干预和设置。

2. 数字通信网及接口，它是数据传输的通道，视频会议系统内的各个会议点之间必须要通过它来实现互连和多媒体信息交换。这要求它的信息速率必须足够高，国外目前随着ISDN的大量建成，已具备了这种条件，而国内仍有很大差距。

视频会议系统的主要功能前面我们已经介绍过了，这里不再赘述，下面只想再举两个典型的视频会议系统的应用例子。

据报道美国已经利用视频会议系统来为小城镇地区的心脏病患者提供及时的X光检查的专家服务，从而加快了抢救病人的速度，提高了诊断的正确性。另外美国密苏里乡村也已经利用视频会议系统为原来无法请到专职教师的学校提供授课服务。由此可见，视频会议系统不仅可应用于一般的视频会议，而且还可在许多应用领域发挥作用，就像上面提到的远程医疗服务、远程教学，乃至远程购物、储蓄、购票等等。

三、个人可视通信

个人可视电话通信一直是人们的梦想，因为人们习惯于面对面交谈，不仅要听到对方的“音容”，还要看到对方的“笑貌”。现在随着宽带通信网(如ISDN)的建立和视频、音频压缩/还原技术的进步，这个梦想可望不久之后得到实现，因为国际电信组织已在积极考虑可视电话的实施问题，而相关的ISDN标准和视频通信标准也已经出台，所以可视电话正在逐步变为现实。

个人可视通信又简称个人视频，它不同于“房间对房间”的视频会议系统和多媒体会议系统，它的基本着眼点在于“面对面”的个人通信，其视频终端机的应用环境将面向普通办公场所；另外个人视频也不像视频会议系统那样正式，令人不舒适，它被设计得很温馨、友好，使用起来也很自然方便，可以说是一

个十分亲密的伙伴。

个人视频除用于个人通信之外还可应用于商业领域，例如：房地产经纪人可以与未来的购买者在线上共同观看有关房间情况的图画；而银行也可以把“活”的视频“出纳员”派到远地点的客户计算机屏幕上，完成客户的存款、取款、转账等业务。显然这一新型业务也可同样应用于保险公司、旅行社、服装公司等，实现与客户的异地交易。

第四节　多媒体地理信息系统

对于地理信息系统，大家也许比较陌生。地理信息系统，顾名思义，就是与地理位置有关的信息处理系统，如铁路调度系统、铁路计划统计管理信息系统、电网管理信息系统、电信综合管理系统等，这些管理信息系统与一般管理信息系统相比，突出地表现在地理位置高度分散这一点上。而一般的管理信息系统尽管也可能分散在各处，但一般地限于一个办公室，一个公司内部。既然地理信息系统与地理位置紧密地联系在一起，所以信息的表现自然也与地图紧密的联系在一起，当然这里的地图意义比较广，比如电信管理系统中由电话局、配线箱、长途电话线等构成的电信网的地域分布也是一种地图；另外由于各信息源在空间上有很大距离，所以数据的传输自然离不开通信网。因

此地理信息系统(GIS,即 Geographic Information System 的首字母缩写)是一项集通信技术、计算机多媒体技术、地图信息管理技术,分布式数据库技术乃至计划、决策技术等多种技术于一体的综合信息管理系统。由于多媒体技术提供了存取、处理各种复杂形式数据的手段,如地理信息系统中涉及到的背景构造,具有图形、图像、声音、空间位置及空间规模的复合信息实体等,所以采用多媒体技术能够高效、直观地实现地理信息系统,从而为地理信息系统的发展开辟了广阔的前景。

不过确切地说,多媒体地理信息系统还只是提供了一个开发多媒体应用地理信息系统的开发平台(或环境),它可以来构造多种应用系统,比如铁道部"八五"重点项目,铁路计划统计管理信息系统(简称 RSMIS)就是这样一个多媒体地理应用系统,它是在地图管理系统 Mapinfo 的环境下开发的。所以在地理信息系统性能的好坏往往也决定着一个应用系统性能的好坏。下面就以 RSMIS 和 MAPINFO 为例介绍一下这方面的情况。

大家知道,铁路部门是一个在地理上分布很分散的系统,其管理系统是一个典型的地理信息系统。事实上其"信息资源中有 85%以上的数据具有地理信息",分析一下 RSMIS 的数据,不难发现,各个子系统的数据都有一个共同点,即同运输设备和运营单位直接或间接相关,运输设备上附着几乎所有各个专业统计信息。以线路为例,它既有线路本身状况(营业长度、单复线、牵引类型等)及运输能力、客货运、机务信息,也有基建、更新改造和大修投资情况,甚至还包括人员劳资信息及施工信息等。而运输设备的运营单位都有在地图上相对应的地理对象,如果

能通过地理对象把各个子系统的数据在地图上表示出来，其内含的地理信息便可以一目了然了。

MAPINFO 正是这样的系统，它不仅提供了数据的地图显示能力，还提供了十分强大的地理分析能力。首先 MAPINFO 具有数据分层，对象分层的特点，能满足处理综合信息的要求。于是，一幅矢量地图上可以涵盖 RSMIS 各个专业统计的信息，实现了有效的综合。其次 MAPINFO 采用矢量地图，可以任意放缩，使得从宏观把握到细节深究的转换上灵活自由。目前已经在 RSMIS 系统中实现的 MAPINFO 应用，能实现对象自动标注和查找，不但能对运输和修建指标进行图表分析，直观图分析，还可以改变其地理对象的颜色和大小，对各项指标进行单值的、区域的和一定范围的宏观分析，并直观地显示出各项指标的分布发展趋势。可以说，真正从处理纷繁的报表文件而陷入的“只见树木，不见森林”或“只见森林，不见树木”的困境中解放出来了。

郑州铁路局作为 RSMIS 的试点单位，目前已建立起铁道部、郑州铁路局及所属六个分局的三级分布式数据库，并完成了运输计划统计，投资计划统计两个子系统的应用开发，投入试运行，并已于 1995 年 12 月 28 日，通过了铁道部中期评审，已开始在全路推广。

RSMIS 系统的建成和使用，无疑可以提高铁路运营信息采集的实时性，通过计算机辅助分类和综合，可以更好更快地获取铁路运营状况，从而更好地掌握铁路经营发展和变化的规律，有利于及时做出科学计划和决策，指导铁路运输工作，提高铁路利

用效率。铁路计划管理信息系统的初步建成和投入使用，也为我国多媒体地理信息系统的应用开创了一个新局面，同时也向铁路管理现代化迈出了坚实的一步。

第五节　多媒体监控系统

记得小时候看过一部科幻小说，说的是发生一起凶杀案后，为识别出是谁作的案，侦探搬来了一台红外热像成像仪，能把几小时前案犯留下的身影摄下来。其根据是人体是个红外线源且人体的体温一般都高于室温，而且人体温度是个缓变场，所以即使一二小时前留下的“温度身影”仍可用热像成像仪摄录下来。现在到底能不能根据这个原理制造出这种成像仪我不太清楚，不过我确实知道实用的夜视监控系统已经问世了，它能工作在漆黑的环境下，这就是多媒体夜视监控系统。

多媒体夜视监控系统在硬件上可分为三个部分，即监视头、总控制台和多媒体计算机控制中心。监视头也称前端，它直接安放在被监控对象的现场，由夜视摄像机、探测器及监听器等组成，构成对现场的声、像信息的实时采集；不同现场的前端采集到的多路数据汇集在一起被送到总控制台，然后再分发到多媒体计算机控制中心的相应设备上：图像显示在显示器上或存储起来，声音可以通过扬声器播放出来或录下来，甚至和报警装置

连在一起，产生触发报警的信息。

夜视监控系统可在停电及漆黑的环境下工作，能监听、监视，并可自动摄像，录下盗贼在现场作案的真实情况，自动记录下作案时间，自动进行报警，为主人进行安全防患准备条件，而且还可增加无线传输、无线报警、无线遥控、信号加密、编解码等新的功能，甚至有可能的话还可以与公安报警系统的报警装置连在一起，实现自动报案。

多媒体夜视监控系统做到了对现场的实时监视和监听，自动录音、摄像等功能，提高了监控系统的自动化程度，也增加了保安人员的安全性，甚至能实现无人监控，该系统可广泛应用于银行、机关、商店、公安、监狱、看守所、重要仓库等的安全监控。

青少年百科

QINGSHAONIAN BAIKE

责任编辑：马 俊 许国萍 李群英
责任校对：李永锋
封面设计：碧 云

ISBN 978-7-5373-1083-3